2007

Optimizing the sensitivity and specificity of genetic disease detection / by Travis E. Marion.

Marion, Travis Emerson

Optimizing the Sensitivity and Specificity of Genetic Disease Detection

By

Travis E. Marion

Masters Degree in Science Thesis

Presented to the Department of Biology

Your file Votre référence
ISBN: 978-0-494-33580-2
Our file Notre référence
ISBN: 978-0-494-33580-2

DECLARATION

I hereby declare that the work presented in this Masters Thesis is original, except where otherwise acknowledged and submitted to fulfill the requirements of a Masters of Biology at Lakehead University.

Travis E. Marion

ACKNOWLEDGMENTS

I would like to gratefully acknowledge the supervision of Dr. Carney Matheson. Without his support, guidance, expertise, drive, patience, and above all friendship the planning, study, and writing of this thesis would not be possible.

I would like to acknowledge my graduate committee members and external reviewer for their time and constructive criticisms towards the final revisions.

I wish to extend my thanks to the staff of the Paleo-DNA Laboratory for their help with reagent preparation, training, troubleshooting, and sequencing.

I would like to thank my fellow Masters candidates Ryan Lamers, Shana Hayter, Louise Cloutier, Lucia Iachetta, and Theresa-Lynne Jones for their advice, patience, company, and friendship.

I would like to thank my friends and family for their love and support. Thank you to my parents for instilling me with ambition and teaching me the value of hard work and perseverance. Finally, I would especially like to thank my wife Susan for her love, encouragement, understanding, and patience exhibited towards me throughout this process.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

A	Adenine
aDNA	Ancient DNA
AGE	Agarose Gel Electrophoresis
bp	Base pair
BSA	Bovine serum albumin
C	Cytosine
°C	Degrees Celsius
chDNA	Chloroplastic DNA
cm	Centimetre
ddATP	Dideoxyadenine
ddCTP	Dideoxycytosine
ddGTP	Dideoxyguanine
ddH_2O	Double distilled water
ddNTP	Dideoxynucleotide
ddTTP	Dideoxythymine
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	deoxynucleotide
dsDNA	Double-stranded DNA
DTT	Dithiothreitol
E. coli	*Escherichia coli*
EtBr	Ethidium Bromide
Exo I	Exonuclease I
G	Guanine
g	Gram
Hb	Haemoglobin
Hb C	Haemoglobin C
Hb D	Haemoglobin D
Hb E	Haemoglobin E
Hb O	Haemoglobin O
Hb S	Haemoglobin S (Sickle Cell)
kDa	Kilo-Daltons
M	Molar
mg	Milligram
Mg^{2+}	Magnesium ion
$MgCl_2$	Magnesium chloride
mRNA	Messenger RNA
mtDNA	Mitochondrial DNA
nDNA	Nuclear DNA
O_2	Oxygen
PCR	Polymerase chain reaction
Prot K	Proteinase K
RFLP	Restriction Fragment Length Polymorphism
RNA	Ribonucleic acid
rpm	Rotations per minute
SAP	Shrimp alkaline phosphatase
SDS	Sodium dodecyl sulfate
SNE	Single nucleotide extension
SNP	Single Nucleotide Polymorphism
ssDNA	Single-stranded DNA
STR	Short tandem repeat
T	Thymine
T_a	Annealing temperature
T_m	Melting temperature
Taq	Thermicus aquaticus DNA polymerase
TNE	
µg	Microgram
µL	Microlitre
µM	Micromolar
UV	ultraviolet

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

The molecular diagnosis of genetic disease is of great importance to the field of medicine and medical research. Though many rare inherited diseases have been described in the medical literature, the inheritance pattern of more common diseases has yet to be established. Until recently evolutionary disease research has been conducted either by molecular diagnosis of those living with the disease or by archaeological investigations examining the morphological pathologies of tissue remains. It was the purpose of this study to design and optimize a PCR based multi-stepped multiplexed SNE methodology to detect β-haemoglobinopathy biomarkers that could be applied to degraded medical and archaeological specimens and expand upon previous work. In order to apply the methodology to degraded tissue samples it was hypothesized that the necessary increase specificity and sensitivity could be obtained by increasing the number of primers used to amplify target sequences and by using the products of previous PCR reactions in subsequent PCR reactions respectively. Furthermore by multiplexing and incorporating an SBE SNP detection methodology the amount of time, cost, involved in genetic disease detection could be reduced. The optimal methodology to be followed in future applications consists of three sequential steps: multiplex PCR, hemi-nested PCR, and single base primer extension SNP detection. It was through this methodology that multiple gene amplifications were produced and sequenced from dilution extracts of 1:1000000 that exhibited 100% homology to the reference sequence. With the increase in detection sensitivity and specificity, multiplex capabilities and the involvement of SNP detection this methodology can be expanded to detect multiple genetic mutations/variants and be applied to medical screening, association studies, population mapping, identification of individuals, evolutionary disease studies, and validate pre-existing research.

CHAPTER 1: INTRODUCTION

GENETIC DISEASES

A genetic disease, or disorder, is the result of chromosomal or gene abnormalities/mutations present at the time of conception. Inherited genetic diseases are the emphasis of this study, whereby intact defective haploid gene(s) are passed to offspring from one or both parents. They can exist as single gene disorders where the disorder is the result of a single gene variant passed to subsequent generations. They can be inherited by one of the following inheritance patterns: autosomal dominant, autosomal, recessive, X-linked dominant, X-linked recessive, Y-linked, or mitochondrial. Genetic disorders may be complex whereby they are inherited multifactorially or polygenically and they cannot thus be definitively described.

Genetic diseases are common in all races, sexes, and populations of people. Though their individual frequency is low, together they affect hundreds of millions of people. Depending on the genotypic mutation, genetic diseases can be phenotypically clinically silent or result in 100% mortality. The diagnosis, treatment and prognosis are all dependent on an individual's DNA sequence. Hence, there is an increasing demand for new, fast, reliable, low cost, high throughput nucleic acid based genetic disease detection methodologies to not only detect but understand the genetic mechanisms of inherited diseases.

STRUCTURE OF DNA

Deoxyribonucleic acid (DNA) is an organic polymer found within most organisms. It serves as the building blocks of life containing instructions needed to construct components of cells. In addition it contains the genetic instructions that determine the development and function of every cell within an organism. The nucleus, mitochondria, and chloroplasts are, for the most part, the

cellular organelles that contain DNA in eukaryotes. The combination of nuclear DNA (nDNA), mitochondrial DNA (mtDNA) and chloroplastic DNA (chDNA) found in photosynthetic organisms code all genetically determined traits ranging from metabolism to immunogenic characteristics (Hummel, 2003).

DNA is composed of three biochemically distinct components: nitrogenous bases, phosphates, and sugars, together known as a nucleotide. The structure of eukaryotic DNA is two helical chains (double stranded) coiled around the same axis, each chain consisting of phosphate diester groups joining β-D-dioxyribofuranose residues with 3' and 5' linkages (Watson and Crick, 1953b). These phosphates and dioxyribose sugars, making up the backbone of the molecule, are asymmetrical and run in opposite directions in an anti-parallel arrangement. This structure is a constant among all forms of DNA. The chains are stabilized by the hydrogen bonds that exist between the pairs of nitrogenous bases which can be in four types found as pairs in perpendicular planes to the helical axis (Watson and Crick, 1953a). The nitrogenous bases are as follows: Adenine (A), and guanine (G) are purines and cytosine (C) and thymine (T) are pyrimidines. The combination of these four nitrogenous bases determines the precise function and coding capacity of the DNA (Luftig M, 2001). It is the sequence in which these bases exist that varies, leading to the different genes and alleles leading to the generation of genetic variation. The nitrogenous bases bind to each other in a complimentary pattern known as base-pairing. G and C bind to each other through the formation of three hydrogen bonds while A and T bind to each other through the formation of two hydrogen bonds.

The overall structure of the DNA polymer is a pair of templates in a double helix formation with a right directional rotation as a result of the combined forces phosphodiester bonds, hydrogen bonds, hydrophobic interactions and pi stacking (Watson and Crick, 1953a; Watson

and Crick, 1953b). As the chains wind around each other areas of nitrogenous base pairs are exposed between the gaps of the phosphate sugar backbone. These gaps are known as the major and minor grooves and serve as access points for protein interaction, proteins such as helicase, topoisomerase, and transcription factors.

Nuclear DNA is described as above, a double stranded double helical structure. The haploid genome is composed of up to 3 billion nucleotides long in humans (Kirk et al., 2002). In order for the entire human nuclear genome to fit within the confined space of the nucleus the nDNA is distributed amongst thirteen pairs of compacted DNA/protein complexes known as chromosomes (Luftig M, 2001). The chromosomes are initially formed as 166 base pairs (bp) of the double helical DNA polymer wind around an octamer of proteins known as histones. The histones serve as an organizational structure that provides a region for transcription. The histone/DNA complex is termed nucleosome and as a whole. Each nucleosome is joined to one another by a stretch of linker DNA of 48bp and a single histone protein. Without further coiling the polymer is classified as chromatin (Alberts B, 1994). These nucleosomes and linker strands further coil upon each other six per complete turn and together are referred as a solenoid. The solenoids further supercoil and can be classified as chromosomes.

Nuclear genetics are inherited through Mendelian genetics or the transmission of a double set of 23 homologous chromosomes in the diploid status whereby each set is inherited through the combination of one haploid gamete from each parent (Hummel, 2003). It is however the variations and small changes that occur within the linear sequence of DNA comprising chromosomes that causes random variations leading to no identical sequence of bases between individuals with the exception of some twins (Luftig M, 2001). Thus nuclear DNA is a great method of comparison and identification. As this study is of medical, forensic and

anthropologic/archaeological nature chDNA will not be discussed. Though mtDNA does have the fore-mentioned capabilities the genetic diseases to be discussed are of nuclear origin and similarly mitochondrial DNA with also be excluded.

GENETIC MUTATIONS

Genetic mutations are changes in the nucleotide sequence within a polymer of nucleic acid, either DNA or ribonucleic acid (RNA). Mutations occur during cell division and under the exposure and influence of chemical mutagens, ionizing radiation, ultraviolet radiation, and viruses (Lindahl, 1993; Lindahl and Andersson, 1972; Lindahl and Nyberg, 1972; Mitchell et al., 2005). Some processes, such as cross linking during meiosis, are intentional and propagate genetic diversity. Neutral mutations exist by which the overall fitness of an individual is unaltered. Similarly the majority of mutations in humans are insignificant as our cells have mechanisms that proof read, repair, or dispose of mutations.

A change in genetic information may have varying effects on health dependent on their structural nature and whether they affect the structure or function of essential proteins. Mutations in multicellular eukaryotes can occur in somatic or germ cells. It is only those DNA mutations that produce a phenotypic change or remain within and individual in a carrier state that occur in germ cells during gametogenesis that can be inherited by offspring and are thus important in inherited genetic disease detection studies (Hartl, 2005). Mutations such as these can occur as large scale mutations that affect the structure of the chromosome or as small scale that affect the nucleotide sequence of a small number of genes.

Chromosomal structural mutations can occur as a result of amplification, deletions, translocations, insertions, inversions or loss of heterozygosity. Since these are large scale,

mutations such as these in human rarely result in birth and if so usually result in failure to thrive decreasing the inheritance of the mutation (Hartl, 2005). Amplifications are mutations that duplicate a number of genes that result in multiple copies of chromosomal regions and possible an overall increase in genetic expression. Deletions occur when a region of genetic material is lost leading to the absence of corresponding genetic expression. Translocation mutations occur when genetic material is exchanged from one non-homologous chromosome to another. Insertions occur when genetic material from one chromosome is added to another. Inversions occur when one region is flipped in orientation. This is only significant if the split occurs in a gene or a promoter region. If this occurs it may be silent or be a duplicate if the corresponding material on the homologous chromosome is also passed to the same gamete. Finally loss of heterozygosity occurs when an allele is lost by a recombinant event or a deletion in a cell that previously had two alleles.

Small scale insertion, deletion, and point mutations are more commonly the cause of inherited genetic mutations as they occur more frequently and generally reduce the fitness of an individual to a lesser degree (Hartl, 2005). Insertion mutations occur when one or more nucleotides are added to the sequence of DNA. They usually occur in humans as a result of replication errors during replication of regions containing repeating elements (Alberts, 1994). Translation occurs by reading three nucleotide complement codons of messenger RNA (mRNA), the addition of extra nucleotides results in a frameshift and the production of an altered protein. Insertion mutations may also alter the coding region of a gene and thus cause improper splicing of the transcribed mRNA also resulting in incorrect translation (Weatherall, 1976). However insertions at the small scale can be repaired in some cases by cellular repair mechanisms that excise the inserted nucleotides. Deletions are similar to insertion mutations in that they too can

result in improper splicing of mRNA or a frameshift mutation however these are not reversible by innate repair mechanisms (Collins, 1984). Point mutations occur as a result of a malfunction during the replication of DNA replication that results in the exchange of a single nucleotide for another, known as a single nucleotide polymorphism (SNP) (Brookes, 1999). They can occur as transitions where a purine is exchanged for a purine or a pyrimidine is exchange for a pyrimidine or as transversions where a purine is exchanged for a pyrimidine or vice versa (Alberts, 1994). They are reversible if another point mutation occurs at the same position and the mutation is reversed to the original sequence. They tend to exist throughout the genome however those that are present in protein coding regions can be one of three classifications dependent on the resulting erroneous codon. Silent mutations are point mutations that code for the same amino acid and thus generally occur as the last nucleotide in the codon. A missense mutation codes for a different amino acid. This may have varying degrees of severity on the overall structure and function of the resultant protein and thus fitness of an individual (Hartl, 2005). If the exchanged amino acid possesses similar side chains and thus similar properties then the protein structure may be altered to a lesser degree (Alberts, 1994). If a missense mutation occurs then the original coded amino acid is replaced by a stop codon and results in a prematurely terminated protein (Collins, 1984).

SINGLE NUCLEOTIDE POLYMORPHISMS (SNPs)

Single nucleotide polymorphisms (SNPs) are an abundant form of genomic variation and are of great importance to research crossing multiple disciplines and thus the development of their detection techniques are at the forefront of methodological research. Between individuals 99.9% of our genome is identical, and of that 0.1% variation approximately 90% of which is attributed

to SNPs (Brookes, 1999). They are therefore thought to be responsible for our individual defining characteristics. They are defined in humans as di-allelic single base pair positions in genomic DNA at which different sequence alternatives (alleles) exist in normal individuals in some population(s), wherein the least frequent allele has an abundance of 1% or greater (Brookes, 1999). On average they occur throughout the genome at a frequency of 1/1000 nucleotides (Li and Sadler, 1991; Wang et al., 1998). However the actual distribution of SNPs can exist at frequencies below 0.1% in some coding regions or as high as 10% in some non-coding regions (Guillaudeux et al., 1998; Horton et al., 1998; Nachman et al., 1998). The summation of these numbers results in several million base pair differences between two individual genomes and approximately 100 000 amino acid differences between their proteomes (Brookes, 1999).

SNPs have forensic and paternity/maternity implications. They have the potential to replace existing markers as they are abundantly variant yet serve as very stable genetic markers and have relatively low mutation rates in comparison with other types of genetic markers (Carlson et al., 2001). SNPs are expected to replace existing short tandem repeat identification systems as they are more abundant and are thus more discriminating. Once the variants of SNPs within the human genome have been mapped it is believed that their association with phenotypes will provide an accurate tool to differential between different populations and ethnicities to be used in criminal profiling techniques.

SNPs are expected to take the place of short tandem repeat (STR) polymorphisms as markers in disease gene mapping, just as microsatellite markers replaced restriction fragment length polymorphism (RFLP) markers (Landegren et al., 1998).

In terms of global human population genetics and anthropologic studies SNPs can be used in an evolutionary context to map the divergence of humans and chimpanzees and even the emergence of different populations. Appropriate sets of SNP markers can be exploited much the same as mitochondrial haplotypes have been used previously or used as markers to track movement of ancient populations and confirm the descent of the human race. High throughput cost effective detection systems have the potential to map all common human gene variants and thus establish SNP databases (Landegren et al., 1998). Known SNPs have been used in reverse genetic studies to compose ethnic backgrounds and comprehensive genetic portraits (Faerman et al., 2000).

The human genome project has made it apparent that nDNA SNPs play key roles in determining phenotypes (Larcombe et al., 2005). Most SNPs are not responsible for disease states as only 3-5% of the human genome codes for the production of proteins. However the SNPs located in genes can be expected or directly affect protein structure or expression levels and therefore represent candidate alterations for genetic mechanisms in disease and are therefore perfect candidates for disease detection studies (Landegren et al., 1998). Furthermore the specificity and sensitivity of SNP detection techniques make it possible to analyze archaeological specimens for the presence of diseases attributed by SNPs in hopes of connecting disease research with the field of Anthropology. The evolution of disease and its impacts on ancient peoples can now be traced. Analysis of cytokine SNPs of ancient people has been shown to have applications in understanding differential population health as the human immune system has evolved to allow for human survival despite exposure to infectious environmental agents.

The expanding number of known SNPs, along with greatly improved methods to monitor them, now promises to offer fresh insights into human biology and pathophysiology (Landegren

et al., 1998). They are more prevalent in the genome than microsatellites, furnishing large sets of markers near or in any locus of interest (Landegren et al., 1998). Those SNPs found outside coding regions can serve as biological markers, much the same as RFLPs have been used previously, to pinpoint disease on the human genome map as they are commonly located near a gene associate with a disease state. Clearly many clinical phenotypes do seem to have a genetic component. It is suspected that most common diseases are polygenic, meaning that the variation in the presence/absence of the disease is attributable, at least in part, to polymorphisms of multiple interacting genes (Kirk et al., 2002). Diseases such as cancer, cardiovascular disease, mental illness, autoimmune states, and diabetes, are expected to be heavily influenced by the patterns of SNPs one possesses in certain key susceptibility genes yet to be identified (Brookes, 1999). Furthermore many overlapping alleles, presence of SNPs plus environmental factors may predispose or cause a disease state in an individual. If a factor such as those presented above contributes to an increase risk or causes a disease it should be found at a higher frequency in individuals with that disease compared to non-diseased. Ultimately, studies into disease genetics are trying to determine precisely which genomic sequence variants alter function and thus have a pathogenic effect (Brookes, 1999). The goal of the research community is to develop assays for all human SNPs and score these in large sets of patients plus matched non-diseased controls for all the complex phenotypes one wished to understand.

MOLECULAR DIAGNOSIS

Molecular diagnostics is a branch of clinical diagnostics based upon nucleic acid amplification that uses primarily DNA or RNA originated either from patients or pathogens as the biomarker for clinical testing (Zhao, 2005). The development of molecular diagnostic techniques is of paramount importance as they have made advances in the fields of Medicine,

anthropology, and to some extent forensics. Importantly, they can minimize time and cost as diagnostic tests can be performed directly on clinical specimens without the need for culture. The specificity and sensitivity of diagnostic techniques allows amplification and detection of previously undetected nucleic acids. They can now predict an individual's potential risk for onset or screen for a specific cancer, be used to screen for disease, genetic or infectious, or be used for prenatal testing. In addition because of the genetic distinction of each individual, molecular diagnostic techniques can be used for forensic or identity testing.

Molecular diagnostics combines various gene based amplifications technologies with highly sophisticated detection methods for the clinical diagnosis of a broad range of diseases including infectious diseases, cancer detection, genetic diseases as well as forensic/archaeological identification and population or evolution studies (Faerman et al., 2000; Zhao, 2005). However the most encouraging achievements of molecular genetics in these fields such as personal identification of missing, victims of mass disasters, and historical investigations, clinical, evolution or otherwise, that require the analysis of degraded DNA in archived, forensic, and/or archaeological degraded specimens is still limited by methodological difficulties (Faerman et al., 2000). Furthermore despite the dramatic success of identifying rare Mendelian patterns of inherited disease the previous techniques have yet to identify the genetic factors responsible for common diseases (Carlson et al., 2001). This along with the growing knowledge of the predisposition role of genetic polymorphisms associated with disease risk factors has prompted the need for faster, more reliable, and cost effective genotyping methods for genetic identification of mutations (Xu et al., 2005). Within the scope of clinical diagnosis the emergence and implementation of new technologies of increasing sensitivities and specificities will allow continual improvement of patient care and thus save more lives (Zhao, 2005).

It can be stated that the purpose of this study is to to design and optimize multiple molecular genetic disease detection systems of increasing specificity and sensitivity on modern extracts that can be applied to degraded tissue samples. The amplification process or gene based technologies to be optimized, though discussed in detail later, typically consists of template hybridization, nucleic acid synthesis. There are generally three classes of gene-based amplification technologies: 1) target based, which are designed to detect and amplify the target gene of interest; 2) probe based, which rely on amplification of the probes that are homologous to a specific gene target; and 3) signal based technologies that amplify the signal rather than the gene target sequence (Zhao, 2005). Though each of these technologies possesses unique advantages and all have the common goal of simplifying and expediting mutation analysis, the nature of this study will only concentrate target based technologies and more specifically polymerase chain reaction methods.

CHAPTER 2: DEGRADED DNA

DEGRADED AND ANCIENT DNA

The purpose of this study is to develop a genetic disease detection methodology that can not only be applied high quality extracts but is sensitive and specific enough that it can be used to investigate the presence of multiple diseases in degraded medically archived samples and archaeological specimens. In doing so this methodology can be used to conduct multiple tests on preserved patient samples, confirm correct diagnosis of previously medically archived specimens, detect the presence and role of disease in ancient populations, and validate any pre-existing theories pertaining to diseases contributing to morphological changes in ancient tissue remains.

The development of the PCR has made it possible to analyze ancient and degraded DNA sequences. The terms ancient and degraded will be used interchangeably throughout this section as a number of references apply to both conditions and fail to distinguish them from each other. It allows the amplification of minute amounts of DNA in the presence of large amounts of damaged and degraded DNA as well as specific amplification in the presence of vast amounts of irrelevant environmental DNA that may remain in extracts (Handt et al., 1994). This advantage has permitted researchers to explore the genetic structure of ancient samples in hope of answering the questions associated with the origin of humankind, the advent of ancient societies, the construction of social organizations, migration patterns, expansions, extinctions, and the spread and evolution of both genetic and infectious disease (Cipollaro et al., 2005). The study of degraded DNA has now matured into a research tool relevant to disciplines as diverse as forensic science, archaeology, conservation biology, taxonomy, zoology, all medical fields.

When DNA is extracted from ancient/degraded remains a number of obstacles must be considered. It is generally found that only a small fraction of specimens contain endogenous, ancient DNA (aDNA) sequences that can be amplified by the Polymerase Chain Reaction (PCR) (Hoss et al., 1996), thus suggesting the necessity for increased PCR cycles. Of those sample extracts that can be amplified the fragments are limited to a low molecular size of 40-500bp (Paabo, 1989) and primers should therefore be designed to amplify fragments in this range. The likely reason for this reduction in fragment size is damage and decay to the DNA that occurs over time and is characterized by strand breaks, abasic sites, miscoding lesions, and crosslinks (Willerslev and Cooper, 2005). This damage can be attributed to spontaneous degradation, and/or environmental exposure, such as low level radiation, or genotoxic chemicals (DNA damaging agents), and/or extraction/purification processes (Hofreiter et al., 2001; Mitchell et al., 2005). This damage will progressively accumulate until the DNA molecule loses its integrity and decomposes, with an irreversible loss of nucleotide sequence information (Paabo et al., 2004). Therefore the occurrence of postmortem damage in DNA extracted from archaeological specimens is a serious concern pertinent to the study of ancient DNA as it may make the application of many molecular biological techniques difficult or impossible, and/or cause erroneous sequence information to be obtained (Higuchi et al., 1987; Paabo et al., 2004).

DNA DEGRADATION

DNA is a relatively unstable biological molecule. It is continually affected by destructive chemical forces (Handt et al., 1994; Paabo et al., 2004). However within the living cell, the integrity of the DNA molecule is protected by cellular membranes and is continually being maintained by various host enzymatic repair pathways (Lindahl, 1993). In contrast to these processes, *in vitro* samples or when a cell becomes metabolically inactive (dead or dormant)

DNA exhibits instability and may be spontaneously hydrolyzed, oxidized or degraded by endogenous nucleases (Hofreiter et al., 2001; Mitchell et al., 2005; Paabo et al., 2004; Willerslev and Cooper, 2005) as the cellular compartments that normal sequester catabolic enzyme breakdown (Hofreiter et al., 2001). In addition to these processes, insects, bacteria, and fungi that feed on macromolecules may be present and accelerate DNA degradation in postmortem samples (Paabo et al., 2004).

Intrinsic destabilization of a nucleotide base is a predominant method of degradation experienced by the DNA molecule (Mitchell et al., 2005). The instability and rate of decay of its primary structure has been investigated and determined to favour depurination at a rate 20 times faster that the loss of pyrimidines (Lindahl and Nyberg, 1972). Instability can cause the cleavage of the phosphate sugar backbone, or single strand break, as well as the loss of a nucleotide base (Mitchell et al., 2005). Furthermore at the site in which a nucleotide base has been lost the aldehyde form of the deoxyribose sugar is rendered vulnerable to cleavage by β-elimination. The end product is a degraded sugar residue and DNA fragments with terminal 3' and 5' phosphate groups.

DNA is a highly hydrophilic molecule and therefore remains hydrated even in arid climates (Handt et al., 1994). It is therefore apparent that in addition to molecular instability, hydrolysis can too result in the loss of a nitrogenous base and formation of strand breaks (single-stranded nicks) (Mitchell et al., 2005; Paabo et al., 2004). Single strand breaks can occur by direct cleavage of the phosphodiester backbone of the molecule (Lindahl, 1993; Willerslev and Cooper, 2005). Alternatively, base loss, usually depurination, is more likely to occur as the glycosidic bonds between the nucleotide base and the sugar are the main target of hydrolytic cleavage (Poinar, 2003) (Lindahl and Nyberg, 1972; Willerslev and Cooper, 2005). After a site has lost its

base, the sugar, may undergo a conformational change to its aldehyde form in which in a few days will undergo β-elimination cleavage and strand breakage by alkali or DNA specific endonucleases (Lindahl, 1993; Lindahl and Andersson, 1972). Purines are released from DNA molecules at a similar rate (guanine released only slightly more rapidly), whereas cytosine and thymine are lost at a 5% higher rate (Lindahl, 1993). Single stranded DNA depurinates at a rate four times that of double stranded DNA as the double helical structure provides only a minimal increase in protection (Lindahl, 1993).

Hydrolytic forces additionally have the ability to deaminate, or displace amino groups from DNA bases with secondary amino groups, and thus create a miscoding modification (Poinar, 2003) (Lindahl, 1993; Paabo et al., 2004; Willerslev and Cooper, 2005). The conversion of cytosine and its homologue 5-methylcytosine to uracil and thymine respectively are the most common deamination reactions (Lindahl, 1993). This reaction involves direct deamination by alkali-catalyzed hydrolysis or the direct attack by water on the protonated base in a general acid catalysed reaction and leads to C-G to T-A transversions (Cipollaro et al., 2005; Lindahl, 1993). Similarly deamination of adenine and guanine may occur at reduced rates resulting in their conversion to hypoxanthine and xanthine respectively (Lindahl, 1993; Paabo et al., 2004). In contrast to hydrolytic base loss, the double helix DNA structure offers good protection against hydrolytic cytosine deamination occurring at a rate 0.5-0.7% that of single-stranded DNA molecules (Lindahl, 1993; Mitchell et al., 2005). These conversions are noteworthy as they cause *Taq* DNA polymerase to insert incorrect bases during the PCR and produce erroneous amplifications (Willerslev and Cooper, 2005).

The amplification of a target DNA sequence may be limited due to the oxidative forces that have been subjected to it as these forces can result in major helical distortion (Lindahl, 1993).

Free radicals such as peroxide radicals, hydrogen peroxide, and hydroxyl radicals can induce lesions that block *Taq* DNA polymerase's ability to extend DNA strands and thus inhibit PCR (Hofreiter et al., 2001; Paabo et al., 2004). In addition to these lesions the double bonds of both purines and pyrimidines, as well as the 3-4' carbon bond of the deoxyribose are sites that are highly susceptible to oxidative attack and result in ring fragmentations (Paabo et al., 2004). These purine and pyrimidine sites of oxidative attack may lead to bases modification. Nine different pyrimidine and purine oxidized modifications have been identified (Hoss et al., 1996). Of these modified bases, those of pyrimidines, in particular thymine, are known to be occur more frequently as they are more sensitive to oxidative damage (Teoule in Paabo, 1989). Pyrimidines are oxidatively modified to form blocking lesions known as 5-hydroxy-5-methylhydantoin and 5-hydroxyhydantoin (Paabo et al., 2004) whose presence inversely affects one's ability to amplify aDNA sequences (Hoss et al., 1996). Hydroxyl radicals commonly generate 8-hydroxyguanine, a base lesion that base-pairs preferentially with adenine rather than cytosine (Lindahl, 1993).

Crosslinking is another type of DNA damage that can block the actions of *Taq* DNA polymerase and hinder PCR amplification (Paabo et al., 2004). Crosslinks may occur intramolecularly or intermolecularly (Willerslev and Cooper, 2005). Intramolecular or intrastrand crosslinks arise as a result of alkylation between nucleotide bases (Willerslev and Cooper, 2005). The presence of Maillard products can allow for the formation of intermolecular crosslinks (Willerslev and Cooper, 2005). Maillard products are the result of condensation reactions between sugars and the primary amino groups of nucleic acids (Paabo et al., 2004).

Molecular instability, hydrolysis, oxidation, deamination, and crosslinking are processes that can contribute to the degradation of the DNA molecule over time and jeopardize its integrity. As

time passes, the cumulative effects of damage to the DNA will be so extensive that no useful molecules remain (Hofreiter et al., 2001). However, when considering certain processes, no correlation exists between the degree of fragmentation and the age of a specimen and it is generally agreed that the state of DNA preservation is not necessarily related to the age of the specimen (Faerman et al., 1995; Hagelberg et al., 1991; Paabo, 1989). This may be due to the fact that said processes will reach a plateau after which they only have minimal effects (Paabo, 1989). Alternatively the environmental conditions and temperature in which the DNA samples have been preserved, rather than age are of decisive importance when considering the integrity of a samples sequence (Paabo, 1989). The environmental and preservation conditions are not always static nor do they impose the same destructive forces upon the DNA molecule. For instance, under circumstances, such as rapid desiccation, low temperatures of high salt concentration, nucleases can become destroyed or inactivated before all nucleic acids are reduced to mononucleotides (Hofreiter et al., 2001). Thus the post-mortem desiccation of a sample can reduce DNA degradation. It is therefore apparent that the taphonomy of an aDNA sample is an important factor when considering DNA degradation.

CONTAMINATION

When attempting to amplify and sequence low copy, ancient and/or degraded DNA, contamination is of utmost concern as it can lead to false positive interpretation of results. If contamination occurs and is in greater, non-fragmented concentrations the contaminant template will more likely hybridize with primers and therefore preferentially amplify in greater numbers than the ancient template (Gilbert, 2005). There are a number of sources of possible contamination. Each require different precautions in order to minimize their risk of occurrence

and the adequate set of control samples and/or control amplifications (Hummel, 2003). These sources can be categorized into contaminations occurring by handling and those contaminated by manufactured products. The former can further be categorized into contaminations occurring by carry-over, cross contamination, and persons handling.

CONTAMINATION THROUGH HANDLING

To fully appreciate the complexities of contamination problems associated with handling one must consider the scales that are involved. Contamination can be an extremely difficult factor to exclude especially if the genuine template is so rare that amplification is possible in only some of the reactions (Hofreiter et al., 2001). One of the most methodological problems to overcome in such analyses is the presence of modern environmental contamination on the surfaces of certain tissues (Kemp and Smith, 2005). Even a few molecules of modern contaminant DNA in extracts prepared from ancient tissues will be amplified and lead to false interpretations of results (Handt et al., 1994). Thus the application of sodium hypochlorite is a technique that can be used to destroy contaminating DNA on the surface of specimens or samples (Kemp and Smith, 2005). A successful, non contaminated PCR can contain $10^{12} - 10^{15}$ amplified molecules in a volume of less than 50μL (Kwok and Higuchi, 1989). Standard procedures such as opening tubes and transferring liquids can create microscopic aerosol droplets that contain 10^6 molecules/0.005 μL of said sample that can easily be distributed on to every surface in or connected to the laboratory via personal movement and/or air exchange systems (Handt et al., 1996; Willerslev and Cooper, 2005). Contamination may additionally occur by any direct contact the pre-PCR sample has with contaminated equipment, gloves or reagents (Herrmann and Hummel, 1994).

This method of contamination whereby previously amplified products are introduced into a pre-PCR step is known as carry-over contamination (Hummel, 2003). If this type of contamination occurs all future amplifications using the same reagents may reveal the exact same genotype of the introduced amplification product (Hummel, 2003). Carry-over contamination however can be prevented by physically separating pre and post-PCR workspaces, adopting one-way traffic, and eliminating the exchange of equipment or reagents between workspaces (Herrmann and Hummel, 1994). In addition PCR amplification from trace amounts of contaminating DNA can be suppressed by the ionization of water and formation of free radicals when surfaces are misted with ddH_2O and irradiated whereby the efficiency of decontamination is dependent upon the length of the contaminant DNA sequence, the components of the solution it suspended in, and the length of irradiation (Deragon et al., 1990).

Cross contamination in an alternative source of contamination in which sample to sample contamination occurs (Hummel, 2003). To minimize this method of contamination, it is necessary to take special consideration when performing all pre-PCR procedures (Hummel, 2003).

It can be assumed that personnel that work or have previously worked in the laboratory or have handled samples may have contaminated them. When working with samples of human or even Neanderthal origin it is wise to use individual specific markers, and have present and past personnel sequences in a database in order to identify contamination events (Hummel, 2003) if sample can be spared. However if an ancient sample is severely contaminated due to handling of unknown persons, it unfortunately cannot be determined through the use of negative controls sample nor by the reproduction of results in secondary facilities (Hummel, 2003). Only if the knowledge of the sample's population is known and sufficient knowledge of their mitochondrial

sequences has been accumulated is it possible to predict what sequences can be reasonably expected when applied to certain mitochondrial analysis (Handt et al., 1994). An added difficulty exist when amplifying microbial DNA remains as many papers have suggested that an exceptional similarity between sequenced ancient microbial samples and the sequences of contemporary ones exists (Graur and Pupko, 2001). The complete exclusion of re-colonized ancient materials over time is difficult if not impossible to obtain in ancient microbial studies with the passive or active movement of cells in the environment thus supporting the need for independent replication (Willerslev and Cooper, 2005).

CONTAMINATION FROM MANUFACTURED PRODUCTS

Human and microbial DNA and cells are ubiquitous in all laboratory settings (Willerslev and Cooper, 2005). It is therefore prudent to assume that all laboratory reagents, laboratory equipment, and disposables purchased directly from manufacturers are potential sources of contamination that exist beyond the laboratory's influence (Hummel, 2003; Willerslev and Cooper, 2005). Laboratory reagents and commercial equipment marked 'sterile' are not guaranteed to be free of nucleic acids or even viable cells (Willerslev and Cooper, 2005). Products sold as 'DNA free' (or less than 5×10^{-11} mg of DNA) are based on microbiological testing, not PCR (Hummel, 2003). PCR amplification tubes, in particular have been determined to be sources of contamination (Hauswirth, 1994) as human mDNA, control region V, sequences of 121 bp have been detected in up to 90% of tubes tested after 45 amplification cycles (Hummel, 2003). This can be overcome however if amplifying DNA sequences longer than 200-250 bp as autoclaving will fragment contaminant DNA (Hummel, 2003).

AVOIDING CONTAMINATION

With contamination a great concern, strict precautions and controlled conditions exist to avoid contamination with extraneous DNA in PCR (Hofreiter et al., 2001). These strategies exist in attempt to avoid all forms of contamination and thus false positive results. The following is a summary of said summaries, some of which may be a reiteration of the information provided above. Complete physical separation of pre- and post-PCR workspaces, and eliminating the exchange of reagents, and equipment between areas is necessary as is the adoption of a one-way traffic system (Hummel, 2003). Extensive cleaning of reagents, equipment, and workspaces is essential with complete and effective decontamination involving prolonged ultraviolet (UV) irradiation exposure, baking, acid or sodium hypochlorite treatment (Herrmann and Hummel, 1994; Willerslev and Cooper, 2005). Samples should always be kept under sterile conditions and separate from each other (Herrmann and Hummel, 1994). All samples should be handled only while wearing sterile disposable gloves and gloves should be changed between sample handling (Herrmann and Hummel, 1994; Hummel, 2003). Precise pipetting is essential, and performed only with the use of positive displacement pipettes and/or aerosol tight tips (Hummel, 2003). Touching the inside of the tubes with pipette tips should be avoided (Hummel, 2003). Despite these precautions, false-positive reactions may still occur due to low levels of contaminating amplification products that are carried over on the skin, hair, and clothing of the laboratory workers within various laboratories (Rys and Persing, 1993). However by following these steps provided above, one may significantly minimize the risk of contamination and thus false positives.

MONITORING CONTAMINATION

By employing different types of controls the identification of false positives can be achieved the possibility of contamination can be monitored. Negative template controls are subjected to the same essential procedures as the samples being analyzed (Herrmann and Hummel, 1994). A lack of amplification in these samples ensures that the primers used in the amplification are only specific to the sample template. Therefore a negative control is one that exhibits similar physiochemical properties and possesses a similar state of preservation as possible when compared to those under investigation (Herrmann and Hummel, 1994). For instance if an ancient human specific DNA sample is to be investigated from bone extracts, one would choose an animal bone from the same burial site to serve as the negative control (Herrmann and Hummel, 1994). A blind control undergoes all analysis procedures but does not contain any sample material, therefore revealing any contamination in any reagent or chemical used in the procedures (Herrmann and Hummel, 1994). A no-template control or blank serves to detect any amplification contaminants as it contains all PCR reagents except template (Herrmann and Hummel, 1994).

AUTHENTICITY OF RESULTS

Recent aDNA analysis techniques have provided the genetic information of extinct species and populations (Gilbert et al., 2005) that have the power to determine the recent evolutionary history of humans, domesticates, and the pathogens they harboured (Poinar, 2003). However, as previously stated, the amount of amplifiable DNA differs between samples as their survival is dependent on the different conditions in which samples have been preserved. With very little or no DNA surviving in ancient samples these aDNA analysis techniques are subjected to

sensitivity, degradation, and contamination problems that make achieving authentic results difficult (Gilbert et al., 2005; Poinar, 2003).Therefore to increase the likelihood that amplified and sequenced aDNA samples are authentic a set of criteria have been imposed to ensure, with the greatest extent possible, the authenticity of results through meticulous replication and authentication (Poinar, 2003). The criteria can be found in Hofreiter, 2001 (Hofreiter et al., 2001).

The authors argue that in the absence of full compliance to all criteria, the reliability and authenticity of results can be considered uncertain (Gilbert et al., 2005). However few published aDNA studies appear to have adopted all these methods of authentication (Caramelli et al., 2003; Cooper et al., 2001). This may be due to the fact that application of all of these criteria, though designed to ensure the highest quality of aDNA data and conclusions possible, can be unreasonable when considering their application to each different ancient sample type and preservation. It has also been suggested that the authenticity and reliability of aDNA data arises from the complex interplay of several poorly understood areas of knowledge, thus no definite answer exists as to what constitutes a reliable study (Gilbert et al., 2005).

Though these criteria are not to be dismissed, consideration should be given on a case-by-case basis as to whether the evidence presented is strong enough to satisfy authenticity given the problems (Gilbert et al., 2005). For instance one must consider cost and time required for their complete adoption as contamination affects samples differently (Gilbert et al., 2005). Samples may have been previously contaminated by related DNA present in the environment and thus contamination can be considered inevitable. If samples are in fact contaminated prior to analysis these criteria provide no way to detect so. In addition when working with ancient and degraded samples often very little sample exists thus limiting the application of all criteria.

CHAPTER 3: METHODOLOGICAL BACKGROUND

SAMPLES

Genetic disease testing in a medical scenario is a tool used to diagnose the presence or absence of genetic vulnerabilities to disease or to establish ancestry. It is generally used as a diagnostic method in conjunction with an individual's medical history, physical, and any associated medical tests/imaging (Bickley, 2007). Available types of testing include prenatal screening, diagnosis, newborn screening, carrier state testing, predictive and pre-symptomatic testing, research and even forensic testing. With the exception of any post-mortem investigations, the samples to be analyzed are obtained in a manner that attempts to reduce the amount of discomfort to the patient. Commonly pre-mortem tissue samples include blood, hair, skin, and amniotic fluid (Aula and Aula, 2006; Miller et al., 2006). When applied to post mortem tissue, samples may be limited and therefore a choice may not exist, nor is patient discomfort and issue. In all cases the extraction procedure will determine what samples can be analyzed.

EXTRACTION

The extraction process whereby DNA is separated from its host cell or place of deposition is by far the most crucial step in genetic analysis. A mistake at this initial stage may minimize or destroy the potential information preserved and retained in the sample over time (Hummel, 2003). Though less of a concern when working with tissue samples extracted from living donors, optimal DNA recovery during the extraction process is of great importance when working with limited ancient, degraded, or forensic tissue samples as the all subsequent stages of analysis are dependent on the extraction methodology (Hummel, 2003). That being said, there is no one extraction methodology that is optimal for all samples. Rather the method employed is

dependent on the quality, quantity, and preservation method of the sample to be analyzed as well as the tissue type.

POLYMERASE CHAIN REACTION

The polymerase chain reaction (PCR) is a widespread genetic research technique that serves to exponentially amplify a specific genetic sequence until a desired concentration is reached. Since its development, the PCR has been described as having transformed the manner in which scientific research is conducted. The technique is simplistic and its application is supported by a broad variety of commercially available kits. It has proven itself as a valid method of physically separating any particular sequence of interest from its context; and then provides for an *in vitro* amplification of this sequence virtually without limit (Mullis, 1990). The PCR provides an extremely sensitive and relatively straightforward means to amplify very small amounts of DNA or RNA from a single copy of a gene, to milligram amounts of the same sequences consisting of millions or billions of identical copies (Gu, 1995), a process that know takes a matter of hours rather than days.

The procedure, originally described by Dr. Kary Mullis (Mullis et al., 1986), when reduced to its most basic terms simply involves the combination of sample DNA with oligonucleotides primers, deoxynucleotide triphosphates (dNTPs), and DNA polymerase in a suitable buffer followed by repetitively heating and cooling the mixture for several hours until the desired amount of amplification is needed (Erlich, 1989). However one of the initial drawbacks of this procedure was that of the thermolability of the *Escherichia coli* (*E. coli*) DNA polymerase used to catalyze the extension of the annealed primers. Polymerases are necessary as they select the correct nucleotide to add to the primer to extend the DNA chain according to the standard nucleotide base pairing (McPherson, 2000). The always catalyze the synthesis of DNA in the

5'→3' direction (McPherson, 2000). The heat denaturation step, required to separate newly synthesized strands of DNA made the addition of fresh enzyme during each cycle a necessity (Saiki et al., 1988). This problem was overcome through the replacement of *E. coli* DNA polyemerase with a thermostable DNA polymerase purified from the thermophilic bacterium, *Thermicus aquaticus* (*Taq*), that can survive extended incubation at 95 °C (Saiki et al., 1988). Though it has limited chain synthesizing abilities at temperatures >90 °C, the enzyme remains relatively stable to and is not denatured irreversibly by exposure to high temperatures (Erlich, 1989). Subsequent experimentation resulted in a substantial increase in the specificity, efficiency, and possible amplification lengths due to the increased temperature at which primer extension could occur (Saiki et al., 1988).

TAQ DNA POLYMERASE

Taq DNA polymerase is a 94 kDa enzymatic protein with a processivity of 50-60 nucleotides and an approximate extension rate (processivity) of 50-60 nucleotides/second at 70 °C (Innis et al., 1988; McPherson, 2000). It is capable of efficiently amplifying products up to 2-4 kbp. Though it does have 5' → 3' exonuclease proof-reading it does not possess a 3'→5' exonuclease proof-reading activity and thus incorrect nucleotide incorporations can be incorporated. This unfortunately results in decreased accuracy as *Taq* DNA polymerase will yield one base substitution error per 9000 nucleotides and one frameshift error per 41 000 nucleotides polymerized at a temperature of 70 °C (Tindall and Kunkel, 1988). If these errors occur in early cycles a population of incorrect sequence amplifications may be detectable and cause false interpretations.

The *Taq* DNA Polymerase used in this project is Platinum® *Taq* (Invitrogen). This particular *Taq* DNA Polymerase is bound with a proprietary antibody that bocks polymerase activity at

ambient temperatures and can only be activated by a "hot start" (the initial 94 °C denaturation step). This property provides increased sensitivity, specificity, and yield by allowing mastermix assembly at room temperatures without the fear of nonspecific extension or amplification. The manufacturer claims the use of this antibody *Taq* DNA Polymerase complex helps reduce PCR optimization requirements, reaction set-up time, and effort therefore improving PCR results.

PCR METHODOLOGY

The PCR is an *in vitro* method for the enzymatic synthesis of specific DNA sequences using two oligonucleotide primers that hybridize to opposite strands and flank the region of interest in the target DNA (Erlich, 1989). The procedure consists of three essential temperature regimes: template denaturation, primer annealing (hybridization), and extension of annealed primers (elongation) (Mullis et al., 1986).

Template denaturation is the initial phase in which the double stranded template DNA is separated into two complementary single strands. This is typically accomplished by increasing the reaction's temperature to 94 °C for 60 s, with an initial denaturation of 94 °C for 5 min (Harris and Jones, 1997; Hummel, 2003). This amount of high energy input provides the necessary force to break the hydrogen bonds between complementary bases allowing the two strands to separate (Harris and Jones, 1997; Hummel, 2003). Single stranded template DNA is a prerequisite for a successful annealing phase. The annealing phase subsequently follows denaturation whereby primers anneal to their complementary single stranded sequences on the target DNA template. This phase is conducted at an annealing temperature (T_a) which is usually about 5 °C below the true melting temperature (T_m) (Harris and Jones, 1997). A fragment can only be amplified if two oligonucleotides are bound to the ssDNA templates in such a way that

the polymerase extension product of the one contains a binding site for the other (Mullis, 1990). The extension phase is the final step in which a thermostable DNA polymerase synthesizes a complimentary copy of the initial single strand of DNA by incorporating dNTPs to the 3' end of both primers making exact copies of both strands of DNA (Harris and Jones, 1997; Hummel, 2003). These extension products that are created are defined by the 5' end of the primers and indefinite at the 3' ends (Erlich, 1989; Harris and Jones, 1997; Mullis et al., 1986). The annealing and extension phases are also typically conducted for a duration of 60 s if the fragments to be amplified are of less than 2 kb (Hummel, 2003). Two double stranded templates are the resultant. These three phases are subsequently repeated respectively in a series of cycles whereby the products from the preceding reaction cycle serve as the reactants in the next reaction catalyzing a doubling, 2^n, with each cycle, where n is the number of cycles (Erlich, 1989; Mullis, 1990; Saiki et al., 1988). The end result is process in which an inherently specific exponential accumulation of extension products flanked by both primer sets occurs (Harris and Jones, 1997; Mullis et al., 1986).

Since the first report of specific DNA amplification using PCR, the number of different applications has grown steadily, as have the modifications to the basic method. Therefore no single protocol can be considered appropriate, and certainly not optimal for all amplifications (Erlich, 1989; Harris and Jones, 1997). Optimization is always required and requires a delicate balance between the amplification of specific products and the elimination of nonspecific products (Harris and Jones, 1997). For instance, the temperature at which exact denaturation will occur is dependent on the size and base sequence of the target DNA. This phase is of critical importance as the DNA strand will fail to separate with inadequate heating temperature or time. Rather the DNA will remain in its original, more stable double stranded configuration

thus limiting the potential of the annealing phase (Harris and Jones, 1997). Conversely, a denaturation temperature that is too high or persists for an extended period of time may lead to an unnecessary loss of DNA polymerase activity thus reducing the yield of amplified product (Harris and Jones, 1997). A temperature of 94°C will however work well in most cases as 90°-92°C will fully denature any sequence that has a balance GC/AT content (Erlich, 1989). The annealing temperature is a fundamental PCR parameter. It is found to influence both specificity and efficiency of amplification. The annealing temperature is dependent on the length and individual base pair composition of each primer and therefore must be optimized with every new primer set. Lower annealing temperatures result in high efficiency amplifications with low specificity due to mispriming and nonspecific amplification (Chou et al., 1992; Harris and Jones, 1997) whereas the opposite is true of higher annealing temperatures (Harris and Jones, 1997; Hummel, 2003).

PRIMER DESIGN

The design and implementation of PCR primers is of paramount importance as they, more than any other factor, determine the success or failure of an amplification reaction (Erlich, 1989). Therefore the goal of successful primer design is to obtain an optimal balance between achieving specificity and efficiency of amplification (Dieffenbach et al., 1993). Efficiency is defined as how close a primer pair is able to amplify a product to the theoretical optimum (Dieffenbach et al., 1993). It requires optimization of cycling parameters, reagent concentrations and is dependent on the length of both the primers and the fragment to be amplified (Dieffenbach et al., 1993; Harris and Jones, 1997). Both of these criteria are controlled by the intended location of hybridization, primer homology, primer length, amplified fragment length and the annealing temperature of the PCR reaction (Dieffenbach et al., 1993; Erlich, 1989; Hummel, 2003).

The initial step of primer design is to determine the sequence that is to be amplified. Once this is established primers are to be designed to flank the intended target sequence. The specificity of the reaction is ensured when the primers are a complete match with the intended hybridization site, are unique within the genome in focus, and are of appropriate length (Hummel, 2003). Homology can be ensured by determining the primers genetic specificity through a "BLAST" search in GeneBank (Hummel, 2003). One must take the necessary precautions to avoid incorporating possible sequence polymorphisms in to the primer sequence (Hummel, 2003). Oligonucleotides between 18 and 24 bases in length tend to be very sequence specific if the annealing temperature of the PCR reaction is set within a few degrees of the primer T_m (Dieffenbach et al., 1993). The optimal annealing temperature at which primers, under 20 bp in length will hybridize can be calculated as an estimate of T_m, where $T_m = 4(G+C) + 2(A+T)$, whereas when calculating the T_m of longer primers one must, employ computer programs to take into account thermodynamic parameters. The upper limit on primer length is somewhat less critical and has more to do with reaction efficiency (Dieffenbach et al., 1993). Shorter primers will anneal more rapidly to target DNA and form a stable double stranded template but may be less specific (Dieffenbach et al., 1993). Additional bases at the 5' end of primers can be added without affecting the annealing of the sequence specific portion of the primer. These serve as mobility modifiers, sites of restriction digestion, or to eliminating non-specific primer dime accumulation (Brownie et al., 1997; Dieffenbach et al., 1993).

There are a few guidelines to consider when designing PCR primers. For instance the primers should maintain a reasonable GC content as oligonucleotides of 20 bases in length with 50% G+C content have T_m values in the range of 56°-62°C (Dieffenbach et al., 1993). Primer pairs with poorly matched T_m can be less efficient and specific because loss of specificity arises

with a lower T_m and the primer with the lower T_m will have a greater chance of mispriming under these conditions (Dieffenbach et al., 1993). In addition when possible, select primers with a random base distribution and with a GC content similar to the fragment being amplified (Erlich, 1989) and contain approximately equal numbers of each nucleotide. Avoid sequences with significant secondary structure, particularly at the 3' end of the primer (Erlich, 1989). Avoid runs of three or more G or Cs at the 3' end as this can lead to mispriming at GC rich regions (McPherson, 2000). Avoid repetitive sequences or regions containing stretches of the same nucleotide as this can lead to primer slippage on the template (McPherson, 2000). Check the primers against each other for homology particularly at the 3' end which may lead to primer dimer (Erlich, 1989). "Primer dimer" is a double stranded template that is approximately the same length as the sum of both primers and appears to occur when one primer is extended by the polymerase over itself or the other (McPherson, 2000). Hairpin structures caused by self dimerization, and primer dimers impact reaction efficiency and should be avoided as they cause a drastic reduction in the reaction component resources (dNTPs, primers, etc.) (Brownie et al., 1997; Dieffenbach et al., 1993; Hummel, 2003).

If the information provided above is taken into consideration when designing primers, the specificity and efficiency of the reaction can be improved. However due to the complex interactions between various components with each amplification conditions, cycling parametres and reagent concentrations must be established experimentally. Even after optimization one must still consider the template itself, if its integrity is jeopardized the sequence to be amplified may not be completely intact, if present at all.

MULTIPLEX PCR

The multiplex PCR is a modification of the original polymerase chain reaction (PCR). It relies on the same principles, procedures, and role of reagents as the original but differs with the addition of two or more primer sets and therefore can simultaneously amplify two or more specific sequences within a single reaction. Hence by combining more than one primer set the multiplex PCR consumes fewer reagents and offers convenience with regards to expense, time, and effort. The procedure was first described in 1988 (Chamberlain et al., 1988) and has since been successfully applied in many areas of DNA research such as gene deletion analysis (Chamberlain et al., 1988; Henegariu et al., 1994), quantitative analysis (Mansfield et al., 1993; Zimmermann et al., 1996), and analysis of mutations and polymorphism (Mutirangura et al., 1993; Rithidech et al., 1997; Shuber et al., 1993). It has proven to be a valuable tool in the detection of infectious diseases (Harris et al., 1998; Hendolin et al., 1997; Heredia et al., 1996; Markoulatos et al., 1999) and is predicted to aid in the detection and analysis of SNPs in ancient and degraded DNA samples.

Though the multiplex PCR has proven to be a beneficial research and diagnostic tool, similarly to singleplex, no single protocol can be considered appropriate for all amplifications. Multiple primers sets in a multiplex PCR are required to amplify unique target regions as pairs or a combination of pairs under single reaction conditions. It is therefore necessary that all constituents and parameters of newly designed multiplex PCR reactions be optimized empirically as they can possess inherent difficulties. Poor specificity, sensitivity, and/or preferential amplification of certain specific targets, problems associated with singleplex PCRs, may be intensified due to the presence of more than one primer set. Thus the design of a multiplex PCR is a strategic process that requires multiple attempts to optimize conditions (Markoulatos et al., 2003). As a result studies have been conducted in order to determine the

factors in which the design and analysis of a multiplex reaction can be influenced (Henegariu et al., 1997; Markoulatos et al., 1999; Markoulatos et al., 2003; Markoulatos et al., 2002; Vandenvelde et al., 1990).

Design and Optimization

Primer design and implementation, as previously mentioned, is the initial and most important procedure when creating a multiplex. As in a singleplex PCR, the primer sequences determine the target sequence to be amplified, the annealing temperature, and the specificity of hybridization. To ensure a high quality highly specific reaction a number of parameters must be taken into consideration. For optimal hybridization the primer homology for all primers must exhibit 100% complementation to the desired sequence to be amplified and as little complementation to undesired template sequences as possible. Additionally, the annealing temperature of each primer must be at a low enough temperature to ensure primer hybridization yet it must be high enough temperature to ensure no dimerization and specific hybridization of all primer sets. This annealing temperature is usually between 55 and 60 °C and is established experimentally by the use of a gradient PCR (Markoulatos et al., 1999). Each primer must therefore be designed to have similar T_m to the supplementary primers and each primer should not display significant homology either internally or to one another (Henegariu et al., 1997).

The annealing position of primer sets relative to one another is of importance when initially designing a multiplex PCR. If, for instance, two primers of the same sense and equal template affinity are in competition with a single antisense primer, the one sense primer located one closer to the antisense primer has an advantage if the reaction is not limited by too low primer concentration (Repp et al., 1993). In another instance if both sense primers bind to a single

template molecule at the same time, the strand elongation initiated by the 5' primer will stop at the position of the 3' primer, as this primer already blocks the template molecule (Markoulatos et al., 1999).

Optimal primer to template ratio is an essential concept when considering the design of a multiplex PCR. Less than optimal conditions will result in the formation of different sized non-specific amplicons. The optimization of a multiplex PCR should therefore aim to minimize such non-specific interactions. A primer concentration of 10^7 molar excess with respect to template is the standard amount added to a reaction however due to the risk of primer dimer formation the primer concentration cannot be raised significantly higher than 0.5 μm regardless of template concentration (Markoulatos et al., 2002). If the ratio is too high due to a high concentration of primers or a low concentration of viable template the primers will exhibit weak template-independent interactions with each other and give rise to primer dimers. A reaction in which the primer to template ratio is too low does not have the necessary amount of primers to continue synthesizing amplicons from newly amplified products. The resulting effect is the renaturation of denatured strands without further amplification and a reduction in exponential accumulation of product occurs. If fewer than one thousand copies of PCR template are present to be amplified primer dimers and other spurious artifacts caused by primer mispairing may again be preferentially formed thus consuming dNTPs, occupying the enzyme, further reducing desirable product yield, and ultimately complicating genomic information (Ruano et al., 1989; Tindall and Kunkel, 1988).

The concentration of individual primers is of importance and is a reflection of the competitive nature between simultaneous DNA amplifications. Preferential amplification of one target sequence over another is a common occurrence in multiplex PCRs. PCR selection and

PCR drift have been identified as two processes that can influence this event. PCR drift occurs in the presence of low template concentrations and is the result of random events occurring in the early cycles of a reaction (Mutter and Boynton, 1995). PCR selection however is the result of differential primer affinity due to differences in the primary or secondary structure of DNA at potential target sites that favour primer amplification (the target's flanking sequence or entire genome) (Markoulatos et al., 2002; Wagner, 1994). Factors affecting affinity replication rates across template species include different melting temperatures of different primers in the multiplex reaction mix. Different binding constants of different primers at the elongation temperature and secondary structure formation of templates in the annealing stage are the resulting effect, thereby causing stearic hindrance for primer binding (Wagner, 1994). However Primers of nearly identical optimum annealing temperatures theoretically should work under similar conditions if they anneal with single copy sequences. If all the primers in a reaction anneal with equal efficiencies under similar conditions they can all be generally used at the same concentrations (Markoulatos et al., 1999). In determining primer efficiency equimolar primer concentrations (0.1-0.5 μM (Markoulatos et al., 2002)) should be added to the initial multiplex PCR (Henegariu et al., 1997). The results of the initial PCR will suggest how the individual primer concentrations and other parameters need to be changed. Observed unequal amplification can be corrected with an increase in the amount of primers for the "weak" or less intense loci and a decrease in the amount of primers for the "strong" primers or more intense loci. The final concentration of the primers (00.4-0.5 μM) may vary considerably among the loci and is established experimentally (Markoulatos et al., 2002).

A delicate balance exists between dNTP and $MgCl_2$ concentrations within a multiplex PCR reaction as the concentration of the former is dependent on free Mg^{2+}. The optimal dNTP

concentration can be determined by maintaining a constant $MgCl_2$ concentration of 2 mM while increasing dNTP concentrations stepwise from 0.5-1.6 mM (Markoulatos et al., 2002). Concentrations above this value rapidly inhibit amplification whereas concentrations below permit amplification but with reduced amount of product yield (Markoulatos et al., 1999). This is because all PCR amplification reactions must contain free Mg^{2+} in excess of the total dNTP concentration as template DNA primers and Platinum® *Taq* DNA polymerase (Invitrogen) must bind it. Platinum® *Taq* DNA polymerase (Invitrogen) is a magnesium-dependent enzyme that requires an optimal free Mg^{2+} concentration of 0.5 mM for its function (Markoulatos et al., 1999). Optimization of $MgCl_2$ concentration therefore will depend on dNTP, primer and Platinum® *Taq* DNA polymerase (Invitrogen) concentrations. An excess of Mg^{2+}, however, will favour dsDNA and stabilize its structure preventing its complete denaturization (Markoulatos et al., 2002). This will cause a reduction in efficiency and may lead to the stabilization of non-specific primer-template hybridization.

Different buffer concentrations are required for different amplification products. In general, primers sets with longer amplification products work better at lower salt concentrations where longer products become harder to denature (Henegariu et al., 1997). The opposite is true of shorter amplification products. A concentration of 1X is the recommended value for a singleplex PCR. In a multiplex PCR reaction the efficiency can be improved with a 2X increase in buffer concentration (Henegariu et al., 1997; Markoulatos et al., 2002).

Excessive Platinum® *Taq* DNA (Invitrogen) polymerase enzyme quantities can negatively affect a multiplex amplification. Unbalanced amplification at various loci and a slight increase in the back ground may occur possibly due to increased glycerol concentrations in the stock solution (Markoulatos et al., 2002). Experimentation determined that the most efficient

concentrations of *Taq* is approximately 0.4 μL or 2 U/25 μL reaction volume (Henegariu et al., 1997). Conversely the success of the PCR depends on the kinetic advantage that high concentrations of primers have over relatively low concentrations of product. This advantage is lost with insufficient concentrations of enzyme as the equilibrium will persist and DNA strands will rehybridize to each other, displacing annealed primers (Mullis, 1990). However as the extension rates and annealing specificities depends on the availability of various other components and parameters in addition to the enzyme activity, modifications necessary to improve multiplex results must be initially directed to the former components. If difficulty continues to persist in one's multiplex reaction one can significantly improve its performance through the use of adjuvants. Adujants such as DMSO, BSA, glycerol, formamide, and betaine when used in concentrations between 5-10% (vol/vol) improve amplification efficiency and specificity by relaxing the DNA molecule making template denaturation easier (Henegariu et al., 1997; Markoulatos et al., 2002).

The optimization process is long and difficult. It can however be facilitated if attempted in the logical step-wise fashion presented above and summarized below in Figure 1.

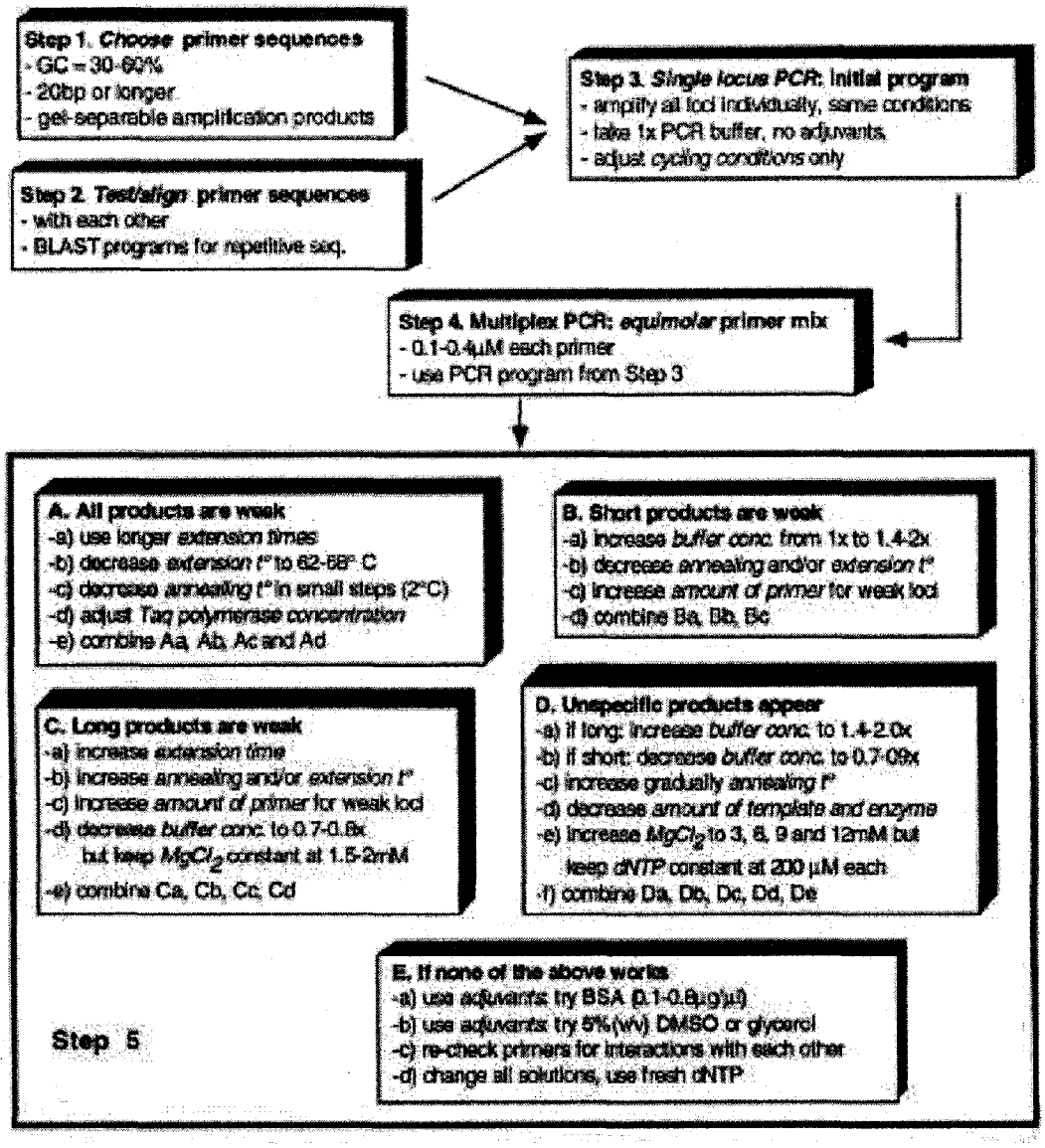


Figure 1: Step-by-Step Multiplex Process
The figure above presents a logical sequential process to experimentally determine the optimal reagent concentrations to be used in newly designed multiplex PCRs (Henegariu et al., 1997).

NESTED PCR

Nonspecific amplification or complete failure to amplify may be the result of the target template sequence. Given the complexity of the human genome and the tendency for many genes to have some sequences in common with other related genes, a lack of specificity is not uncommon in PCR amplifications (Mullis et al., 1986). Alternatively due to the cumulative

effects of damage to the DNA molecule postmortem DNA degradation may result in PCR failure (Hofreiter et al., 2001; Hummel, 2003). aDNA extracts, for instance, do not reveal significant amounts of intact targets exceeding 200bp (Hummel, 2003). The primer hybridization sites or sequences 3' to these locations may simply not be present for amplification resulting in a failure to anneal or extend to the desired length. The nested PCR is a modified variation of the original that may overcome nonspecific amplification or amplification failure.

The nested PCR is a two stage amplification employed to ensure specificity. The initial amplification is a simple PCR that amplifies a locus between the two hybridized oligonucleotides. The second stage of amplification employs one or two oligonucleotides (hemi-nested or nested primers) to prime within the initial amplified fragment to produce a second sub-fragment (Mullis et al., 1986) which may have eluded degradation. Furthermore if a non specific sequence is amplified incorrectly, the probability is very low that it would be amplified again by a second set of primers. This second amplification should lead to a 10^4 enhancement of the desired product over the nonspecific (McPherson, 2000). Thus by employing two successive amplifications coordinated by multiple priming events, one may overcome PCR failure due to degradation as well as increase specificity and yield. This increased specificity provides an easy, direct method of verification.

DNA SEQUENCING

Sequencing is a valuable procedure that enables the analysis of the succession of nucleotide bases within an amplified DNA fragment and thus provides a tool to identify all possible genetic polymorphisms (Hummel, 2003). There are a number of different methods in which sequencing can be accomplished (Hindley, 1983, Wu, 1993) however the method most commonly used is

known as Sanger sequencing and was first described by Sanger (Sanger et al., 1977) and has since been modified for automation.

METHODOLOGY

Sanger sequencing is essentially a single primer PCR performed subsequent to the initial amplification PCR (Hummel, 2003) where the purified initial PCR products serve as the template for the sequencing reaction. This sequencing reaction results in a DNA polymerase-directed synthesis of new DNA from a single primer annealed to a single-stranded template. It initially requires the separation of the preceding purified PCR product's double strands (Sanger et al., 1977) accomplished by a 30 second 94 °C denaturation phase. This is subsequently followed by a 15 second 50 °C annealing phase whereby one primer used to initially amplify the target sequence is hybridized to the PCR product. The 72 °C, 4 minute extension phase is the final step before the cycle repeats its self as many times necessary to achieve the desired concentration of sequence PCR products.

As this sequencing PCR only makes use of one primer the reaction is asymmetric and results in a linear amplification rather than an exponential amplification (Sanger et al., 1977). This is of importance as the goal of this sequencing PCR is to label all base positions with fluorescent 2',3'-dideoxynucleotides (ddNTPs) (Hummel, 2003). This is accomplished during the extension phase where at any base position *Taq* DNA polymerase can incorporate either a 2'-dinucleotide (dNTP) or a ddNTP that corresponds to the template base position (McPherson, 2000). The incorporation of the former results in continued DNA extension as *Taq* DNA polymerase will incorporate one of four dNTPs (either dATP, dGTP, dTTP, or dCTP depending on the corresponding template base) to the 3' end of the previously incorporated dNTP and form a

phosphodiester bond. This newly incorporated dNTP provides the 3' site necessary to add another. The latter however is lacking an additional hydroxyl group (-OH) at the 3' position (Sanger et al., 1977). Therefore the addition of any one of four ddNTPs (either ddATP, ddGTP, ddTTP, or ddCTP depending on the corresponding template base) to a previously incorporated dNTP results in the termination of chain synthesis at its place of incorporation by preventing the formation of the next phosphodiester bond (Sanger et al., 1977) (McPherson, 2000). Provided that the dNTPs and ddNTPs are supplied in suitable proportions only a small portion of these chain termination reactions will occur at random throughout all nucleotide positions of the amplified PCR product (Hummel, 2003). The majority will continue to be synthesized until the next position in which a proportion of the chains will have a terminating ddNTP incorporated (Hummel, 2003; McPherson, 2000). The end result is a generated population of single-stranded products, representing all possible lengths starting from primer plus one to primer $+n$ terminated by one of four fluorescently labeled ddNTPs (Hummel, 2003; McPherson, 2000) compliment to the template. After sample purification, desiccation, and resuspension in a loading buffer the single-stranded labeled sequences can subsequently be sorted by length and detected fluorescently through capillary electrophoresis (McPherson, 2000). The fluorescent signal is digitized and the data is analyzed producing a software generated electropherogram of a successive peaks of different colours representing each of the four ddNTPs (Brandis, 1999; Hummel, 2003; Li et al., 1999; Rosenblum et al., 1997).

The Sanger sequencing method is a rapid and simple method of analyzing sequences. For the methods of interest all the necessary reagents, minus purified amplified sample and primer, are contained in one reagent thus limiting contamination and time. The process does however possess inherent limitations. Sample preparation must be conducted on ice as the reagents may

degrade. The *Taq* DNA polymerase has its own limitations as previously discussed and those resulting in pre-existing PCR errors as well as new ones, may be sequenced and interpreted as valid results.

DETECTION OF SINGLE NUCLEOTIDE POLYMORPHISMS (SNPs)

Comparison of genomic DNA sequences in different individuals has revealed single positions in which two, or sometime more bases can occur (Syvanen, 2001). These single nucleotide polymorphisms (SNPs), as previously stated, are the most widely distributed polymorphisms in the human genome and have emerged as the genetic markers of choice for disease studies due to their high density and relatively even distribution (Chen and Sullivan, 2003; Kruglyak, 1999; Sachidanandam et al., 2001; Venter et al., 2001). Detection and interpretation of SNPs has been recognized as a research tool of enormous potential (Brookes, 1999; Faerman et al., 2000; Filon et al., 1995; Kirk et al., 2002; Larcombe et al., 2005; Risch and Merikangas, 1996; Schork et al., 1998; Wang and Todd, 2003). Although the number of SNP detection methodologies have exploded in recent years and many robust methods are currently available an advancement within the field of Molecular Genetics was required to evolve from labour intensive time consuming and expensive processes to some of the most highly automated, efficient, and relatively inexpensive methods (Kwok and Chen, 2003).

The realization of SNPs potential within the human genome first required the initial study of DNA sequence variation which only became feasible after the discovery of enzymatic and chemical methods of DNA manipulation (Kwok and Chen, 2003). Since then the development of random fragment length polymorphism (RFLP) genome analysis, PCR, perfection and optimization of sequencing techniques, and most notably the human genome project have lead to

their appreciation. As demand for genetic analysis has increased, SNP detection technologies have been developed at an accelerated pace (Kwok and Chen, 2003)

Applications of SNP analyses techniques extend from investigations of small numbers of sequence variants known to be associated with specific disease to investigations of markers across the genome with markers corresponding to several variants of each of the 100000 or so genes (Landegren et al., 1998). Many methods now exist to monitor as little as a single SNP to thousands of SNPs per assay.

These detection techniques encompass two broad areas of research: scanning DNA sequences for previously unknown polymorphisms and screening (genotyping) individuals for known polymorphisms (Kwok and Chen, 2003). The former can further be divided into global, or random, approach and the regional or targeted, approach (Kwok and Chen, 2003). Though the nature of this study and the techniques described and employed emphasize the latter it has the potential to conduct scanning investigations.

With a number of SNP genotyping methods on the market makes the choice of choosing the appropriate method difficult. The ideal SNP genotyping or otherwise detection method must possess the following attributes: 1) The assay must be easily and quickly developed from sequence information; 2) the cost of the assay development must be low in terms of marker specific reagents and time spent by expert personnel on optimization; 3) the reaction must be robust, such that sub-optimal DNA samples will yield reliable results; 4) the assay must be easily automated and must require minimal hands-on operation; 5) the data analysis must be simple, with automated, accurate genotype calling; 6) the reaction format must be flexible and scalable, capable of performing a few hundred to a million assays per day; and 7) once optimized, the total assay cost per genotype must be low (Kwok, 2001). Presently however there is no one ideal

method that is able to meet the needs of all SNP studies (Kwok, 2001; Kwok and Chen, 2003) the aim of this study chose a specific/sensitive methodology that balanced ease for operation and automation with cost.

In general, SNP genotyping protocols consist of three processes. The initial process is target fragment amplification, followed by allelic discrimination and finally by product/identification. A number of methods can be applied for each process, and they can exist in any number of combinations in order to achieve an optimized protocol whereby the gain outweighs the labor and cost (Chen and Sullivan, 2003). The combination methods employed for this study were the PCR for target fragment amplification, primer extension as the method of allelic discrimination, and fluorescence/capillary electrophoresis as the method of allele specific product identification.

The benefits and limitations of using PCR for target fragment amplification have been previously discussed in detail. In summary it works by applying heat to the extracted sample containing mastermix to separate a dsDNA molecule into two single stranded molecules and then copying each of the two single stranded molecules with a thermostable DNA polymerase and an oligonuclotide primer. The end result is two dsDNA copies and when this process is repeated amplification of product is created exponentially. For genotyping purposes target DNA amplification must be specific (Chen and Sullivan, 2003). The sequence amplified must occur at only one locus and pseudogenes cannot exist in other gene families. Once a specific amplicon is determined the PCR, when optimized, ensures specificity by interrogating sequences with a forward and reverse primer of at least 18bp. Its cost and efficiency are of concern. The PCR can be multiplexed to increase efficiency and decrease cost of time and reagents however its success rate can drop to 50-70% and multiple targets may not be amplified equally and the number of samples that can be genotyped (scoring rate) drops with an increase in targets (Chen and

Sullivan, 2003). Furthermore the cost of PCR can be expensive for large scale genotyping studies as licensing fees exist.

Following the PCR step and possible gel electrophoresis a clean-up step is required to remove excess dNTPs and PCR primers left over from the reaction. The clean-up is necessary for all primer extension based methods as the excess dNTPs and PCR primers interfere with primer extension SNP detection methodologies (Chen and Sullivan, 2003). This clean-up process is accomplished enzymatically through the addition of shrimp alkaline phosphatase (SAP) and *Escherichia coli* exonuclease I (Exo I) followed by an incubation step. The incubation process, though it should be optimized consists, of two steps. An initial prolonged incubation at a lower temperature, typically 30 °C for 30 minutes, subsequently followed by a shorter incubation at a higher temperature, typically 95-100 °C for 5 minutes. The initial serves to activate the enzymes and is prolonged to allow enough time for the enzymes to perform their function. The latter is of a high enough temperature and short enough duration to inactivate the enzymes. The SAP is a hydrolase enzyme that functions in an alkaline environment, engineered from a species of arctic shrimp known as *Pandalus borealis*. This enzyme serves to dephosphorylate the phosphate group from the 5' end of nucleotides thus inactivating any remnant dNTPs from reaction products. The exonuclease was first described as a method to rid pre-PCR products of smaller contaminant DNA fragments, and post PCR products of any primer hybridization that could extend upon themselves in subsequent reactions (Zhu et al., 1991). This enzyme serves to catalyze the sequential cleavage of 5' mononucleotides from the hydroxyl end of dsDNA. Thus the two methods compliment each other preventing inappropriate extension and remove any carry-over components of PCR reactions that may inhibit allelic discrimination reactions.

Allelic discrimination technologies are essential genotyping tools. Their individual power of discrimination is dependent on the biochemical mechanism by which they operate (Chen and Sullivan, 2003). The mechanism determines the specificity and accuracy of methods as well as the cost and throughput.

The primer extension is a method of allelic discrimination with a DNA polymerase based mechanism. Though well known, the major function of a DNA polymerase is to replicate DNA during cell division. They are highly accurate and precise, extending an end of a DNA strand and the incoming base match the template exactly (Chen and Sullivan, 2003; Syvanen, 2001). This approach requires amplified target DNA as a template and analyzes the products to determine the identity of the base(s) incorporated at the polymorphic site (Kwok, 2001). It exploits that as DNA polymerases extends DNA strands from 5' → 3' and require a template strand to guide incoming nucleotides the 3' base must be paired with the template (Chen and Sullivan, 2003). However if the 3' extending base is not base-paired to its template DNA the polymerases will not add the unmatched base to the strand. In addition if the 3' extending base is missing the 3-hydroxyl group (-OH), the extension of the strand is terminated at that nucleotide position as the next base cannot be added.

Primer extension allelic discrimination methods are highly flexible and require only one primer set per SNP position. The primer design and optimization is relatively straight forward and the reaction employs the same if not similar reagents to that of the initial PCR step (Chen and Sullivan, 2003; Kwok, 2001; Kwok and Chen, 2003; Syvanen, 2001). Commercially available DNA polymerases such as Platinum® *Taq* DNA polymerase (Invitrogen) do unfortunately exhibit substitution errors. For Platinum® *Taq* DNA polymerase (Invitrogen) this substitution error is one error for every 9000 nucleotides and a frameshift error every 41 000

nucleotides. Although they also have 5' → 3' proofreading ability which is beneficial from a genotyping perspective in order to repair errors and remove oligonucleotide primers used in the DNA replication.

Though numerous variations of the DNA polymerase primer extension method exist the method of interest is categorized as an allele specific primer extension approach (Kwok and Chen, 2003) but is better described as a single-base extension (SBE) (Chen and Sullivan, 2003) or single nucleotide extension (SNE) as it essentially extends an oligonucleotide probe complementary to the target amplification product by one complementary nucleotide. To ensure that the primer extends by only one base the triphosphate nucleotides used in the reaction are ddNTPs and thus do not possess a 3-hydroxyl group (Chen and Sullivan, 2003). The extension primers are designed a kin to PCR primers. They are approximately 20-40 bp in length with ideal annealing temperatures of approximately 50 °C. If multiplexed the extension primers should differ in length by approximately 3-6 bp and exhibit no homology to one another. They are designed to anneal immediately upstream to the target polymorphic site (Chen and Sullivan, 2003). Extension can only occur under the correct temperatures in the presence of a DNA polymerase, extension primers, amplified target DNA, and candidate terminator nucleotides to complement the polymorphic bases. The thermocycling conditions are as follows: denaturation 96 °C for 10 seconds; annealing 50 °C for 15 seconds; and extension 60 °C for 30 seconds, generally repeated 25 times. The polymorphic bases dictate which ddNTP is incorporated. The product is an oligonucleotide that is one base longer than the extension primer. The product is once again cleaned with a SAP treatment and can now be detected.

The single-base extension products are detected by a combination of capillary electrophoresis and fluorescent detection allowing a multiplexed sample to be detected based on those

discriminatory factors. The capillary electrophoresis separates single-base extension products by size. The inserted ddNTP complimentary to the polymorphism is detected by fluorescence. Each of the four terminators is labeled with individual specific fluorophores. The detection is based on excitation of a fluorescent molecule by plane-polarized light, and measurement of the rate of depolarization of fluorescence (Chen and Sullivan, 2003). This rate is proportional to the rate of tumbling of a fluorescent molecule (Syvanen, 2001). Direct fluorescence detection is versatile and can be done in a multiplex to a certain extent (Kwok, 2001). What is observed is a collective measurement of a population of molecules (Chen and Sullivan, 2003). Constants must be maintained as factors such as temperature, viscosity, and nonspecific interactions between fluorescent dyes and other buffer components (Chen and Sullivan, 2003). Thus deviation from a constant environment between samples and reactions may lead to the formation of false positive or false negative results.

CHAPTER 4: PROJECT

HAEMOGLOBIN AND HAEMOGLOBINOPATHIES

Haemoglobinopathies were decided to be the known disease to design and optimize a methodology of increasing sensitivity and specificity that could be applied to the detection of SNPs in this project. To understand why haemoglobinopathies are best suited to model a series of methodologies leading to SNP detection systems one must understand the nature of haemoglobin itself.

HAEMOGLOBIN

Haemoglobin is of fundamental importance to the vital function of erythrocytes. It aids to mediate the exchange of respiratory gases, oxygen and carbon dioxide, between the lungs and tissues (Weatherall, 1976). It allows the transport of one hundred times more oxygen than if oxygen was carried dissolved in plasma alone and as humans, our activity would be restricted to only one fifteenth of our present capabilities (Collins, 1984).

Structurally haemoglobin is a conjugated cytoplasmic protein. It is a spherical, globular molecule with a molecular weight of 64 500 kDa. It is a tetramer consisting of two pairs of polypeptide chains known as alpha (α) and beta (β) globins with a haeme complex attached to each globin (Huehns, 1974). The α-globins consist of 141 amino acids and the β-globins consist of 146 amino acids. Each globin chain consists of eight helical segments (A-H) of nearly identical length and seven non helical segments allowing for bends within the structure. It assumes a structure in which the polar amino acids face the molecular surface and non-polar amino acids face the interior as it is found suspended within the aqueous cytoplasm of erythrocytes (Collins, 1984). The haeme complex of a protoporphyrin ring and an iron complex

and provides a site for oxygen (O_2) to bind. The iron is in the ferrous state and remains so in an oxygenated or deoxygenated state. It is a tetrapyrole, consisting of four pyrole rings. The iron molecule forms bonds with each of the nitrogen atoms on the pyrole rings and with the histidine at globin helix F amino acid position 8 (Collins, 1984). Upon oxygenation the iron binds with the oxygen molecule on the opposite side of the Histidine (His) bond to form an octahedron structure. The oxygen molecule will subsequently bind to the His at helix E position 7(Collins, 1984). This iron/oxygen complex essentially is suspended in a non-polar crevice between the E and F helices. However it forms additional Van der Waal's bonds with many other parts of the molecule adding to the tertiary structure.

The molecule, under normal physiologic conditions, can alter its quaternary structure into one of two forms known as oxygenated and deoxygenated forms (Laberge and Yonetani, 2007). Upon initial oxygenation the molecule transforms its conformation to increase its affinity to reach a point of saturation where the preceding oxygen molecules influence the change. Oxygen's binding affinity is decreased in the presence of carbon dioxide. The products of cellular respiration, carbon dioxide and protons, also cause a conformational change and facilitates the release of oxygen molecules (Eaton et al., 2007). This property amongst others is vital to ensure oxygen retrieval of haemoglobin in oxygen rich environments such as the lungs, and its release in oxygen poor environments such as the body's tissues (Weatherall, 1976).

The genetic mechanisms of the haemoglobin molecule exist on chromosome 16 for the α-globins and chromosome 11 for the β-globins. The alpha and alpha-like genes and beta and beta-like genes exist on the short (*p*) arms of their respective chromosomes (Collins, 1984). Two copies of α-globin structural genes exist on chromosome 16, and are identical except for only one minor difference in the sequence of the second intron. Upstream is a pseudo-alpha gene that

closely resembles the sequence of the alpha genes however a mutation exists within it that renders it non-functional. Further upstream from this are two embryonic like genes, one of which is not expressed. The beta gene complex comprises 6 genes, an embryonic gene, two fetal genes (only one amino acid difference), a pseudo beta gene, δ gene (insignificant), and the beta gene itself on chromosome 11 (Collins, 1984). In both gene clusters the genes are arranged in their order of expression during development. All functional globin genes share a common general structure consisting of three exons and two introns. Both clusters also possess a promoter region approximately 100bp immediately preceding the point of transcription.

HAEMOGLOBINOPATHIES

Haemoglobinopathies can be defined as genetic diseases that result in structurally abnormal haemoglobin molecules (Huehns, 1974). They are caused by mutations in the genetic sequence. Their genetic mechanism, prevalence, severity and impact on global populations make them perfect genetic diseases to design and optimize disease detection methodologies that have the potential to expand to more complex diseases and apply to archived specimens (Faerman et al., 2000; Filon et al., 1995).

As previously described, the genes that code for the α- and β-globin molecules are located on the 16^{th} and 11^{th} chromosome respectively. These regions, unless affected by large scale chromosomal replication or sorting errors are inherited by classic Mendelian patterns of inheritance (Collins, 1984; Weatherall and Clegg, 1976). They can exhibit autosomal co-dominance where symptoms are produced in a heterozygous state. Alternatively they can exhibit a recessive inheritance pattern. Those that are recessively inherited are less likely to be originated as a result of a mutation on the alpha gene as it has two copies from each gamete

whereas there is only one beta gene per haploid genotype. There are over 400 known structural haemoglobin variants (Fucharoen et al., 2002). The variants are prevalent in different populations at different frequencies and may even serve as a tool to track populations (Weatherall and Clegg, 2001). Many genetic variations are either inherited recessively, only occur on one alpha gene, or are just clinically silent. There are over 700 known variants of haemoglobin linked diseases, over 400 of which are attributed to known haemoglobinopathies. 277 of the known haemoglobinopathies occur as a result of a mutation on the beta gene and 242 of which occur as a result of a SNP (Weatherall and Clegg, 2001).

Haemoglobinopathies, whether phenotypically homozygous or genotypically heterozygous have high prevalence in certain regions throughout the world. This can be attributed to the fact that certain heterozygous states are selected for as a protection mechanism of such tropical diseases as malaria (Angel, 1966; Buchanan et al., 2004). The following is a 2001 global summary of approximate numbers of annual births of babies with severe haemoglobin disorders (Weatherall and Clegg, 2001) modified from a 1989 study (WHO, 1989):

Table 1: The frequency of Haemoglobinopathies by region

Region	Population ($X10^6$)	Annual Births ($X10^6$)	Annual Affected Births ($X10^3$)
Sub-Saharan Africa	650	30	230
Americas	730	17.5	5
South-East Asia	3150	84	120
Eastern Mediterranean and Europe	780	11	1.6
Western Pacific	30	0.5	0.2
Total	5340	143.0	356.8

Based on the data presented in Table 1, it is apparent that hundreds of thousands of individuals are estimated to be born with a severe haemoglobinopathy. Of these haemoglobinopathies, severe abnormal haemoglobin S (Hb S) also known as Sickle Cell Anemia, Hb C, Hb D Punjab,

Hb E, and Hb O are known as the common abnormal haemoglobins and reach high frequencies (Fucharoen et al., 2002; Weatherall and Clegg, 2001). Each affects millions of individuals worldwide.

The severity of haemoglobinopathies is dependent on the pattern of inheritance and the mutated gene (Steinberg, 1998). As previously mentioned some may be clinically silent whereas other present as severe haemolytic anemia that will result in eventual death if left untreated (Buchanan et al., 2004). Unfortunately those that are most severe are also most prevalent. Hb S is prevalent particularly in a broad zone of equatorial Africa extending coast to coast and is a cause of haemolytic anemia. The highest incidence occurs in the eastern region of the continent 40-50% of the members of certain tribes are affected in some form (Weatherall and Clegg, 2001). On average however this zone has an incidence of 10-20%. It is also found in non-African populations. A prevalence of 25% has been reported in some areas of Turkey, Saudi Arabia, Israel, and Southern India. It is common in some areas of Greece and along the northern Mediterranean shore. Isolated cases have been observed in Caucasian Americans. Hb C greatest prevalence is found in Western Africa east of the River Niger where 3% of black population carries the Hb C gene either as a heterozygote or homozygote (Livingstone, 1895). Hb E is found most commonly in SE Asia where an estimated 30 million people may carry the gene. Hb D Punjab is found in low percentages India through to the middle-east (Rees et al., 1998).

The severity and prevalence of these diseases has impacted global populations throughout time (Hershkovitz et al., 1997). The diseases are ancient, as they provided genetic protection in the heterozygous state against the prevalence of malaria that followed the establishment of permanent settlements near standing water in the Neolithic period (Filon et al., 1995). This explains their present day prevalence within the equatorial regions.

Haemoglobinopthies are the ideal family of inherited genetic variances to optimize and design detection methods of increasing sensitivity and specificity. It has been stated that they obey classic Mendelian patterns, the cost and time can be reduced as the most prevalent and severe forms are the result of mutations occurring within a few small regions in the beta gene and they are attributed by SNPs. Furthermore when optimized for sensitivity and specificity, due to their relative confinement within the genome, large fragments of preserved DNA are not required. The methodologies developed to analyze the β-globin exons can be applied to any and all medically archived and archaeological specimens from areas with high frequency or prevalence in order to support pre-existing results and develop new theories as to the impact and evolution of the disease.

Until recently the evolutionary study of haemoglobinopathies has been extrapolated from the distribution of present day mutations in different modern populations (Filon et al., 1995) or it has attempted to identify disease in skeletal remains on the basis of bone pathology. The establishment of differential anemias in archaelogical, skeletal and blood specimens remains difficult and many difficulties and methodologic problems in analyzing DNA obtained from archaeological specimens still exist (Faerman et al., 2000; Filon et al., 1995). It is this hope that the developed methodologies will help bridge that gap. Previous studies have been conducted in attempt to already do so. Methodologies have been developed that analyzed bone lesions from skeletal remains that were traced back to Israel during the time of the Ottoman in Empire, suggestive of anemia. DNA analysis confirmed the genetic mechanism causing disease as a mutant haemoglobin allele (Filon et al., 1995). In 2000 a hotstart PCR was developed and successfully amplified DNA from an Jamaican male who deceased in 1993 (Faerman et al., 2000). This study used reverse genetics to compose an ethnic background and a comprehensive

genetic portrait of the deceased individual. The mitochondrial and Y chromosome studies confirmed the individual's West-African descent consistent with the historical account of Africans brought to Jamaica during the trans-Atlantic slave trade. Furthermore the presence of anemic skeletal lesions was confirmed by DNA analysis as the individual was homozygotic HbS.

Through the optimization of newer more advanced detection methods it is the purpose of this study to expand upon its predecessors by developing a multi-stepped, multiplex SNP detection methodology to be applied to degraded tissue samples. It is hypothesized that through subsequent amplifications and primer hybridizations that a methodology of increasing sensitivity and specificity can be obtained that is capable of detecting genetic disease in degraded tissue samples. In doing so the hope of the study is to expand upon previous studies and bridge the gap between the diagnosis and study of β-haemoglobin variants in both modern and ancient specimens (Faerman et al., 2000; Filon et al., 1995). In accomplishing so it is hypothesized that further contribute to the research they have started. Through the multiplex PCRs and SNP detection methodologies this study can more intensely focus on the beta gene cluster while reducing the amount of time, template, reagents, and cost associated with genetic diagnostics. Based on prevalence and frequency of the most severe haemoglobinopathies only those polymorphisms found within the beta gene will be detected.

CHAPTER 5: METHODS

METHODOLOGICAL APPROACH

Figures 2-4 illustrate the stepwise approach taken to design and optimize individual methodologies that once optimized were rearranged in order to provide a final methodology of increasing sensitivity and specificity. A detailed explanation of each optimization procedure can be found under its corresponding heading.

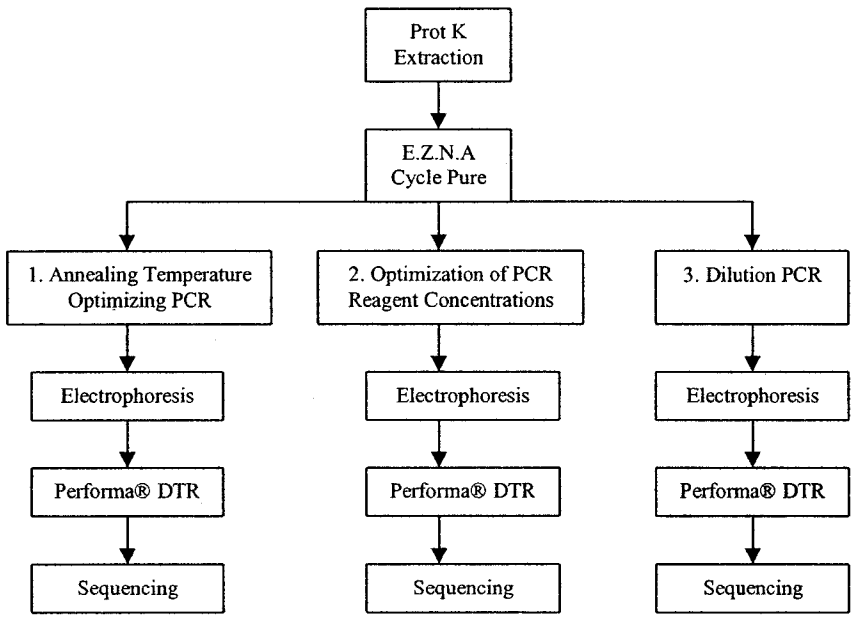


Figure 2: Step 1: Process of Singleplex Optimization of Amplicons I-IV Primer Sets
This procedure was conducted on the primer sets for all amplicon primer sets (Amplicons I-IV). The extract used was the exact same sample for each primer set to ensure continuity of concentration. The protocols are in sequential order listed from top to bottom. 1. The optimization of annealing temperature for each primer sets had to be established first. 2. Once annealing temperatures for each primer set was established the concentrations of each PCR reagent had to be established experimentally for each singleplex primer set. 3. Once annealing temperatures and reagent concentrations for each primer set was established a PCR conducted on a dilution series of the same purified sample using the Amplicon I and II primer sets was conducted to mimic a low copy number template and establish a sensitivity of detection. The products from these amplifications were purified and respectively served as the templates for Amplicon III and IV primer sets. The SBE detection primers were not optimized as singleplexes as the SNaPshot™ kit (Applied Biosystems) all reagent concentrations are to remain the same as indicated in the protocol.

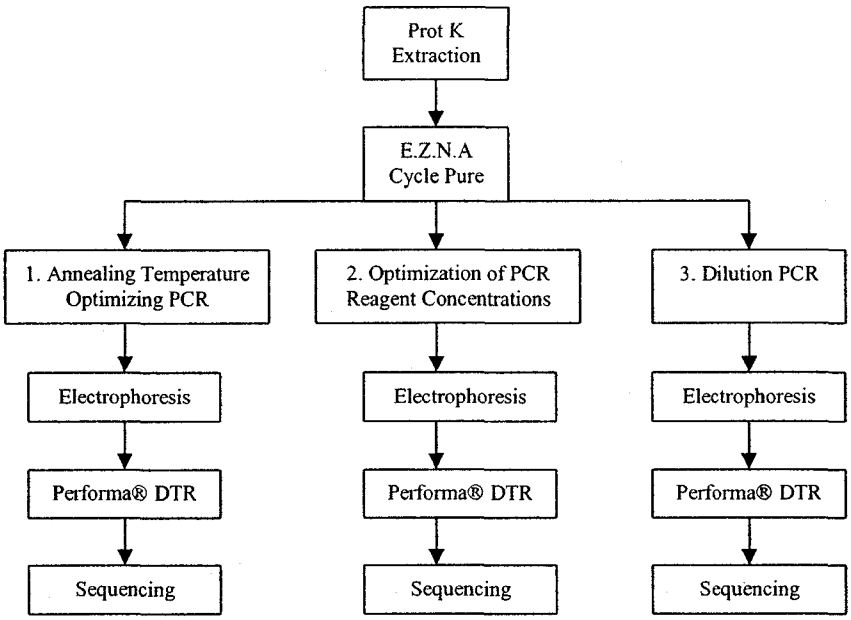


Figure 3: Step2: Process of Optimizing Multiplex and Hemi-nested Primer Sets
With the singleplex optimization of the four primer sets completed, amplicons I and II primer sets could now be optimized as a multiplex PCR and amplicons III and IV could now be optimized as a hemi-nested PCR. The same optimization procedure as in Step 1, Figure 2 was conducted to optimize the multiplex and hemi-nested primer sets. 2. The optimization of PCR reagent concentrations followed the step-by-step process detailed in Figure 1. The SBE detection primers were not optimized as a multiplex as the SNaPshot™ kit (Applied Biosystems) all reagent concentrations are to remain the same as indicated in the protocol and primers were all designed with a similar annealing temperature. Note that the same purified extract was used to optimize both the multiplex and hemi-nested PCRs.

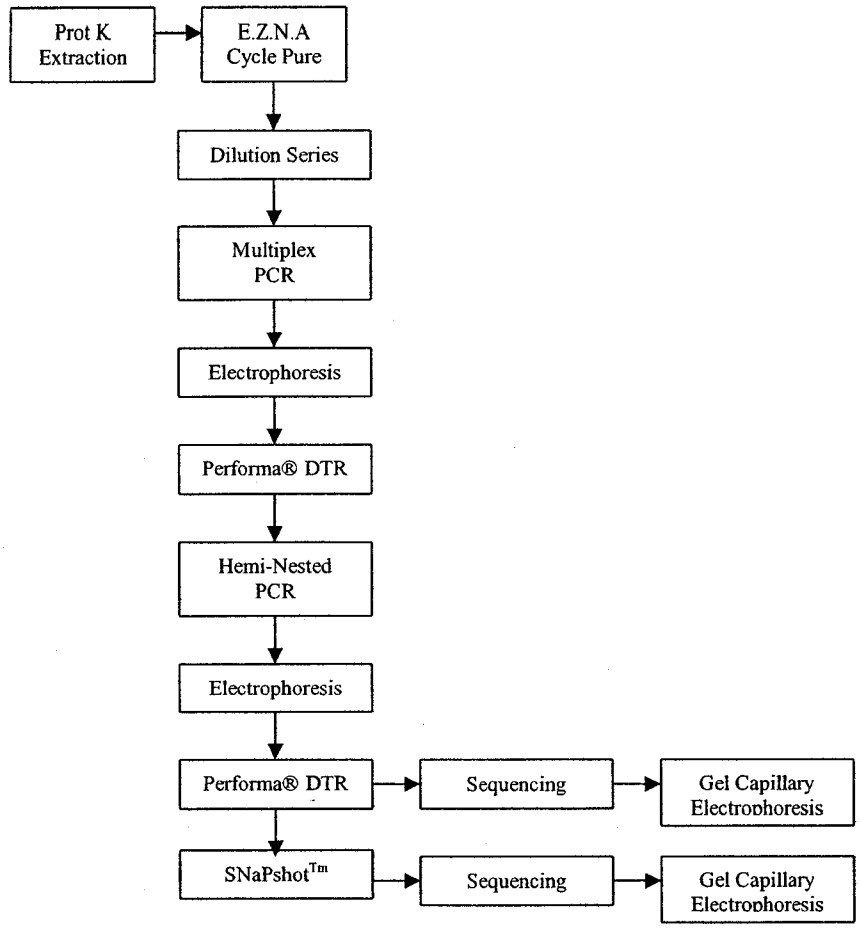


Figure 4: Step 3: Optimized Complete Methodology Applied to Dilution Series.
Once the multiplex, hemi-nested, SNaPshot™ methodologies were optimized individually they were combined in a subsequent fashion. The same purified extract used in the previous steps was used as the initial template in a dilution series (1:10, 1:100, 1:1 000, 1:10 000, 1:100 000 and 1:1 000 000). The dilution series served as the template for the multiplex reaction (amplicon primer sets I and II). The resulting multiplex products (amplicons I and II) were purified and served as the template for the subsequent hemi-nested PCR amplification. The multiplex PCR negative was also purified and used once again as a negative template to ensure no contamination occurred through the purification step. Once purified, the resulting hemi-nested products (amplicons III and IV) served as templates to be sequenced to detect any potential genetic mutations throughout the amplified sequences. The purified hemi-nested products also served as the template for SBE detection of multiple specific, prevalent, and severe β-globinopathies at three nucleotide locations by means of SNaPshot™ (Applied Biosystems).

SAMPLES

The samples tested were the roots of facial hairs, one per extraction and only the bottom 5 mm was used in the extraction. The root was used only if it had a sheath in tact. Due to the time

and cost associated with the ethical clearance of using samples from a population, all samples were modern, self, and considered of high quality.

PROTOCOLS

The following are the protocols used throughout this project. The design and optimization of methods will be discussed in later sections. These are not necessarily presented in order as one can deduce from Figures 2-4. Refer to Appendix A for a list of reagent lot numbers.

EXTRACTION PROTOCOL

The process of extracting genomic material from samples to be used for the optimization of techniques or for future applications was not within the realm of this study. Although it is the most crucial step of any genetic analysis the development of optimal extraction techniques was not attempted, rather it was a project itself completed by a fellow colleague (Hayter, 2007). There simply is no one extraction methodology optimal for all sample types and tophonomies. The method employed is dependent on the quality, quantity, and preservation method of the sample to be analyzed as well as the tissue type. Due to ethical restrictions and in the interest of continuity, subsequent detection techniques were optimized on extracted DNA from self facial hair. Although exact concentrations of nDNA within samples could not be precisely determined all extractions were performed using the identical extraction procedure under as identical environmental conditions as the laboratory could physically permit. Furthermore the same extraction was used to optimize each procedure to ensure continuity. It can therefore be assumed that all samples used in the optimization procedure are of the same quality and quantity.

An optimized proteinase K (prot K) enzymatic method of extraction was used for all self sample extractions (Hayter, 2007; Rohland and Hofreiter, 2007). In the presence of prot K the DNA is released from the cells autolytically (Hansen, 1974). This autolytic system provides a potentially more gentle release of DNA from cells allowing for higher molecular weight fragments to be isolated (Hansen, 1974). The protocol is as follows:

Optimized Proteinase K DNA Extraction (Applied Biosystems)

- Using sterile techniques remove one follicle of facial hair. Ensure that root shealth is present. Cut 0.5 mm from the root and place root into sterile 1.5 mL collection tube.
- Add 290 µL of TNE extraction buffer to sample.
- Add 40 µL of 20% SDS.
- Add 40 µL of 0.39 M DTT.
- Add 2 µL of 20 µg/mL proteinase K.
- Add 28 µL of ddH_2O.
- Incubate tubes at 56 ^{0}C overnight with gentle agitation.

PURIFICATION PROTOCOL

Post extraction the samples underwent a purification process in order to remove unwanted salts, low molecular weight impurities, excess prot K, and to exchange buffers. The E.Z.N.A Cycle-Pure Kit (United Bioinformatica Inc.) was consistently used to purify post-extract samples. It employs a HiBind® matrix that specifically, but reversibly, binds DNA under certain optimal conditions allowing proteins and other contaminants to be washed and removed. The DNA is easily eluted and collected with the use of deionized water or salt buffer. The following procedure was performed on each extracted sample under as similar environmental conditions as the laboratory could physically provide:

E.Z.N.A Cycle Pure Protocol (United Bioinformatica Inc.)

- Transfer to a clean 1.5 mL microfuge tube, and add equal volumes of Buffer XP1. For PCR products <200 bp add 2 volumes of Buffer XP1. Vortex thoroughly to mix.
- Apply the sample to a HiBind® DNA spin-column assembled in a clean 2 mL collection tube (provided) and centrifuge in a microcentrifuge at 10,000 x g (maximum speed) for 1 minute at room temperature. Discard the liquid.

- Wash the column by adding 700 µL of SPW Buffer diluted with absolute ethanol. Centrifuge at 10,000 x g (maximum speed) for 1 minute at room temperature.
- Discard liquid and repeat the previous step with another 700 µL SPW Buffer.
- Discard liquid and centrifuge the empty column for 1 minute at 10,000 x g (maximum speed) to dry the column matrix. This is critical for good DNA yields.
- Place column into a clean 1.5 mL microcentrifuge tube. Add 50 µL sterile deionized water directly onto the column matrix and centrifuge for 1 minute at 10,000 x g (maximum speed) to elute the DNA. This represents approximately 80-90% of bound DNA. An optional second elution will yield any residual DNA, though at a lower concentration though it was never performed.

DETECTION PCR PROTOCOL

Prior to using the extracts in methodological development the purified products the presence of DNA was confirmed by using previously designed primer sets that specifically amplified a portion of the first exon of the β-globin (Beaulne, 2004). All non-primer PCR reagents, detection PCR or otherwise were supplied by Invitrogen. The primers for all reactions were supplied by Operon. This PCR and each subsequent amplification procedure was performed under sterile conditions to prevent contamination, and performed in low light conditions on ice as to not to degrade reagents or purified extracts. The use of ice and Platinum® *Taq* DNA Polymerase (Invitrogen) was employed to prevent undesirable extension product. Furthermore the reactions were mixed in one tube to accommodate for pipetting error. 25 µL of the mastermix were subsequently aliquoted into 0.2 mL tubes. A PCR negative control (water and no DNA) is included to ensure reagent purity. A negative extract control is carried out to test for possible contamination of the extraction procedure. An additional PCR reaction is prepared in case there are pipetting errors or volume preparation of the PCR. The standard concentrations and order of reagents, as well as the Platinum® *Taq* DNA Polymerase (Invitrogen) thermocycler PCR parameters for a single detection PCR amplification are listed below (Table 2):

Table 2: Detection PCR Amplification Final Concentrations

Reagent	Initial Concentration	Final Concentration	µL/Reaction
PCR Buffer	10X	1X	2.5
DNTP mix	10 mM	0.2 mM	0.5
Forward Detection Primer (HBB 3F)	10 µM	0.2 µM	0.5
Reverse Detection Primer (HBB R)	10 µM	0.2 µM	0.5
$MgCl_2$	50 mM	2.0 mM	1.0
Platinum® *Taq*			0.2
ddH_2O			16.8
DNA template			3.0

Table 3: Detection PCR Amplification Parameters

	Temperature (°C)	Time (min)
Hotstart	94.0	2.0
Denature	94.0	1.0
Anneal	63.0	0.5
Extend	72.0	1.5
Hold	4.0	
40 cycles		

DESIGN AND OPTIMIZATION OF MULTIPLEX PCR

Forward and reverse primers were designed to flank the first and second exons of the β-globin gene as the majority of prevalent and severe haemoglobinopathies are confined to these regions. These and all subsequent steps in the complete methodology used Accession M34058 from GENEBank as a reference sequence. The first primer set, HBBIIF (forward) and HBBIIAR (reverse), was designed to amplify a region of DNA 256 bp in length including primer sequences, further referred to as Amplicon I. The second primer set, HBBIIIF (forward) and HBBIIIAR (reverse) was designed to amplify a region of DNA 339 bp in length, further referred to as Amplicon II. Unfortunately due to primer design issues, the second set could not be designed to efficiently flank the second exon. Hence the forward primer is designed within the

exon. Refer to Appendix B for primer set details. The following is a diagrammatic representation of the relationship between the two primer sets, the sequence, and the amplicons, not to scale (Figure 5):

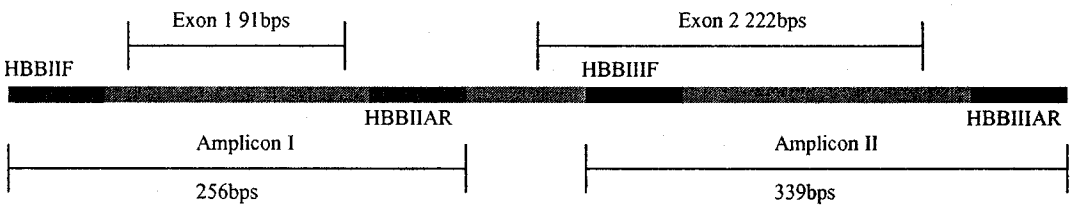


Figure 5: Map of Amplicon I and II Primer Sets and Resulting Amplicons

Each primer set was optimized individually (singleplex) before optimized as a complete multiplex (Figure 2). This optimization required a gradient PCR to establish optimal annealing temperatures. Furthermore titration of all reagents was required to establish the concentrations that would provide the most efficient yet specific amplification. Based on detection PCR of extracts the purified samples are considered to be uncontaminated. A PCR negative control (water and no DNA) is included to each reaction and all subsequent reactions to ensure reagent purity. The PCR preparation was performed as the detection PCR protocol previously described and under similar conditions. The following is the optimized PCR and thermocycling parameters for the individual HBBII and HBBIII primer sets (Tables 4 and 5):

Table 4: Amplicon I Primer Set Amplification

Reagent	Initial Concentration	Final Concentration	µL/Reaction
PCR Buffer	10X	1X	2.5
DNTP mix	10mM	0.2 mM	0.5
Forward Primer (HBBIIF)	10 µM	0.2 µM	0.5
Reverse Primer (HBBIIAR)	10 µM	0.2 µM	0.5
$MgCl_2$	50 mM	1.5 mM	0.75
Platinum® *Taq*			0.2
ddH_2O			17.05
DNA template			3.0

Table 5: Amplicon I Primer Set Thermocycling Parameters

	Temperature (°C)	Time (min)
Hotstart	94.0	2.0
Denature	94.0	1.0
Anneal	63.0	1.0
Extend	72.0	2.0
Hold	4.0	
40 cycles		

Table 6: Amplicon II Primer Set Amplification

Reagent	Initial Concentration	Final Concentration	µL/Reaction
PCR Buffer	10X	1X	2.5
DNTP mix	10 mM	0.2 mM	0.5
Forward Primer (HBBIIIF)	10 µM	0.12 µM	0.3
Reverse Primer (HBBIIIAR)	10 µM	0.12 µM	0.3
$MgCl_2$	50 mM	2.0 mM	1.0
Platinum® *Taq*			0.1
ddH_2O			17.3
DNA template			3.0

Table 7: Amplicon II Primer Set Thermocycling Parameters

	Temperature (°C)	Time (min)
Hotstart	94.0	2.0
Denature	94.0	1.0
Anneal	65.0	1.0
Extend	72.0	1.75
Hold	4.0	
40 cycles		

A dilution series of 1:10, 1:100, 1:1 000, 1:10 000, and 1:100 000 was applied to the optimized protocols amplified and sequenced in order to obtain a sample that not only behaved like low copy number DNA but also would serve as an initial measurement of sensitivity to compare the sensitivity of subsequent methodologies. The Amplicon II primers were additionally applied to a 1:1 000 000 dilution. Each PCR product, optimized and diluted, was subsequently applied to AGE, Performa® DTR Gel Filtration Cartridges (Edge Biosystems) and sequenced according to the described protocols described later to ensure desired amplicon length and specific amplification.

Upon confirmation of specific amplification, the annealing temperature, titrated reagents, and dilution series, the multiplex reaction was conducted and similarly optimized (Figure 3). The following is the optimized PCR and thermocycling parameters for the multiplex reaction using Amplicon I and Amplicon II primer sets (Tables 8 and 9):

Table 8: Multiplex Primer Set Amplification

Reagent	Initial Concentration	Final Concentration	µL/Reaction
PCR Buffer	10X	1X	2.5
DNTP mix	10 mM	0.2 mM	0.5
Forward Primer (HBBIIF)	10 µM	0.12 µM	0.5
Reverse Primer (HBBIIAR)	10 µM	0.12 µM	0.5
Forward Primer (HBBIIIF)	10 µM	0.08 µM	0.2
Reverse Primer (HBBIIIAR)	10 µM	0.08 µM	0.2
$MgCl_2$	50 mM	2.0 mM	1.0
Platinum® *Taq*			0.2
ddH_2O			16.4
DNA template			3.0

Table 9: Multiplex Primer Set Thermocycling Parameters

	Temperature (°C)	Time (min)
Hotstart	94.0	2.0
Denature	94.0	1.0
Anneal	63.0	1.0
Extend	72.0	1.5
Hold	4.0	
40 cycles		

DESIGN AND OPTIMIZATION OF HEMI-NESTED MULTIPLEX PCR

Hemi-nested primers were designed to anneal within the purified, non-sequenced, multiplex amplification products. Refer to Appendix B for primer set details. The first hemi-nested primer, HBB1F, is a forward primer that was designed to anneal to the purified multiplex amplification product Amplicon I, downstream from the original HBBIIF. It in conjunction with HBBIIAR is to use the multiplex amplification products as a template for the next amplification step and production of a 227 bp fragment further referred to as Amplicon III. The second hemi-nested primer, HBBIIINR, is a reverse primer that was designed to anneal to the purified multiplex amplification product Amplicon II upstream from the original HBBIIIAR. It in conjunction with HBBIIIF is to use the multiplex amplification products as a template for the next amplification step and production of another 234 bp fragment further referred to as Amplicon IV. The following is a diagramatic representation of the relationship between the multiplex and the hemi-nested primers, the sequence, and the amplicons, not to scale (Figure 6):

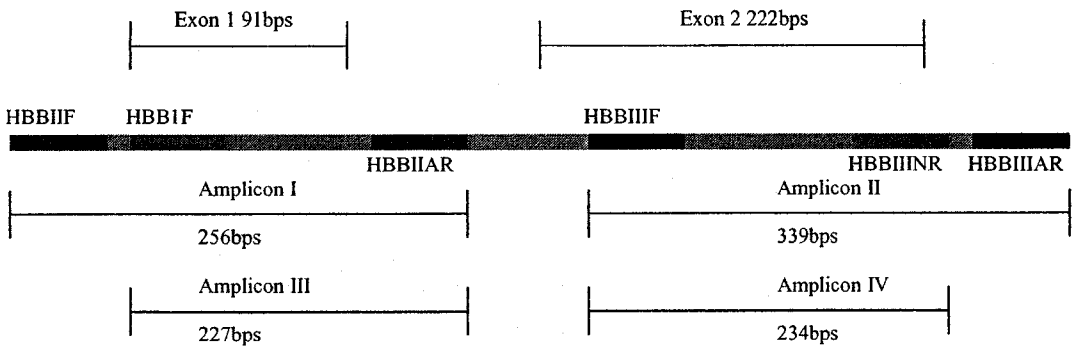


Figure 6: Map of All Amplicon Primer Sets and Resulting Amplicons

Each primer set, HBB1F with HBBIIAR for Amplicon III and HBBIIIF with HBBIIINR for Amplicon IV, was optimized individually (singleplex) before optimized as a complete multiplex (Figure 2). This optimization required a gradient PCR establish optimal annealing temperatures. Furthermore, titration of all reagents was conducted experimentally to establish the concentrations that would provide the most efficient yet specific amplification. Based on detection PCR of extracts the purified samples are considered to be uncontaminated. A PCR negative control (water and no DNA) is included to each reaction and all subsequent reactions to ensure reagent purity. The PCR preparation was performed as the detection PCR protocol previously described and under similar conditions. The following is the optimized PCR and thermocycling parameters for the singleplex Amplicon III and Amplicon IV primer sets (Tables 10 -13):

Table 10: Amplicon III Primer Set Amplification

Reagent	Initial Concentration	Final Concentration	µL/Reaction
PCR Buffer	10X	1X	2.5
DNTP mix	10 mM	0.2 mM	0.5
Forward Detection Primer (HBB1F)	10 µM	0.2 µM	0.5
Reverse Detection Primer (HBBIIAR)	10 µM	0.2 µM	0.5
$MgCl_2$	50 mM	2.0 mM	1.0
Platinum® *Taq*			0.2
ddH_2O			116.8
DNA template			3.0

Table 11: Amplicon III Primer Set Thermocycling Parameters

	Temperature (°C)	Time (min)
Hotstart	94.0	2.0
Denature	94.0	1.0
Anneal	60.0	1.0
Extend	72.0	2.0
Hold	4.0	
40 cycles		

Table 12: Amplicon IV Primer Set Amplification

Reagent	Initial Concentration	Final Concentration	µL/Reaction
PCR Buffer	10X	1X	2.5
DNTP mix	10 mM	0.2 mM	0.5
Forward Detection Primer (HBBIIIF)	10 µM	0.12 µM	0.3
Reverse Detection Primer (HBBIIINR)	10 µM	0.12 µM	0.3
$MgCl_2$	50 mM	2.0 mM	1.0
Platinum® *Taq*			0.1
ddH_2O			17.3
DNA template			3.0

Table 13: Amplicon IV Primer Set Thermocycling Parameters

	Temperature (°C)	Time (min)
Hotstart	94.0	2.0
Denature	94.0	1.0
Anneal	65.0	1.0
Extend	72.0	1.75
Hold	4.0	
40 cycles		

In order to increase sensitivity and specificity, the dilution series purified products from Amplicon I primer set and Amplicon II primer set were used as templates for Amplicon III and IV primer sets respectively (Figure 3). The Amplicon III primer set was used to further amplify the Amplicon I primer set 1:10 000 and 1:100 000 dilution products in order to assess if an increase in specificity could be obtained. The Amplicon IV primer set was used to further amplify the Amplicon II primer set 1:100, 1:1 000, 1:10 000, 1:100 000, and 1:1 000 000 dilution products. The PCR negatives used in each dilution series for the singleplex of Amplicons I and II were purified, carried over and had all reagents added to serve as purification negatives. PCR negatives were also included for both reactions to ensure the purity of reagents used. Each PCR product, optimized and diluted, was subsequently applied to AGE, Performa® DTR Gel Filtration Cartridges (Edge Biosystems) and sequenced according to the previously described protocols to ensure desired amplicon length and specific amplification. Thus the singleplex optimization (Figure 2) of the hemi-nested PCR primer sets III and IV was identical to that of primer sets I and II, except that for the dilution products from I and II served as the templates for primer sets III and IV respectively.

Upon confirmation of specific amplification, the annealing temperature, titrated reagents, and dilution series, of the singleplex hemi-nested primer sets the hemi-nested primer sets were multiplexed and optimized using the same optimization procedure as the multiplex optimization (Figure 3). Once again the purified products from the multiplex dilution served as the hemi-nested multiplex reaction templates to assess the increased level of sensitivity (Figure 4). The PCR negative from the multiplex dilution series was once again carried over, purified, and applied to PCR reagents in order to serve as a purification negative. The following is the

optimized PCR and thermocycling parameters for the hemi-nested multiplex reaction using Amplicon III and Amplicon IV primer sets (Tables 14 and 15):

Table 14: Hemi-Nested Multiplex Primer Set Amplification

Reagent	Initial Concentration	Final Concentration	µL/Reaction
PCR Buffer	10X	1X	2.5
DNTP mix	10 mM	0.2 mM	0.5
Forward Primer (HBB1F)	10 µM	0.16 µM	0.4
Reverse Primer (HBBIIAR)	10 µM	0.16 µM	0.4
Forward Primer (HBBIIIF)	10 µM	0.08 µM	0.2
Reverse Primer (HBBIIINR)	10 µM	0.08 µM	0.2
$MgCl_2$	50 mM	1.5 mM	0.75
Platinum® *Taq*			0.2
ddH_2O			16.85
DNA template			3.0

Table 15: Hemi-Nested Multiplex Primer Set Thermocycling Parameters

	Temperature (°C)	Time (min)
Hotstart	94.0	2.0
Denature	94.0	1.0
Anneal	60.0	1.0
Extend	72.0	1.5
Hold	4.0	
40 cycles		

ELECTROPHORESIS PROTOCOL

The detection and all subsequent PCR products are applied to 2% agarose gel electrophoresis (AGE) containing ethidium bromide (EtBr) and are viewed on a transluminator with ultraviolet (UV) light. The 2% agarose gel preparation and electrophoresis loading/running procedures are listed below:

Agarose Gel Electrophoresis Preparation

- Measure the inside of the plate with a ruler to obtain the length and width.

  L * W * 0.3 = Volume of 1XTBE to use
- Grams of agarose required = (% / 100) * (Volume of 1XTBE)
- Suggested Agarose Concentrations:

Table 16: The fragment sizes best analysed with each Agarose concentration

Size range (base pairs)	*Final agarose concentration (%)*
150-800	1.8
100-600	2.0
50-250	3.0
20-130	4.0
<80	5.0

- Prepare in clean flask.
- Add calculated amount of 1XTBE and agarose (swirl flask to avoid clumping). **Note, if you are using a small boat, use 25 mL of 1XTBE and 0.375 g of agarose to give a 1.5% gel.
- Cover flask with a polystyrene cup in which small holes have been punctured for ventilation.
- Microwave on defrost for 90 seconds (or until agarose dissolves).
- Allow gel solution to cool down (2-4 min.).
- Add 2 µL Ethidium bromide (4 µL for large gel boats).
- Pour the gel into the horizontal apparatus, removing all the bubbles that will affect DNA migration.
- Insert the appropriate size comb to form wells in gel.
- Allow the gel to set for ~ 45 min. at room temperature.

Loading/Running Agarose Gel Electrophoresis

- Fill out a gel loading sheet.
- Turn gel sideways in apparatus.
- Pour 1XTBE buffer in electrophoretic apparatus until gel is covered (~ 1 cm).
- Remove comb carefully.
- Purge wells with transfer pipette.
- Load wells with sample and loading dye, and one lane of molecular marker.
- Set voltage 110 V and time at ~ 30 min. on electrophoretic apparatus.
- After gel has run, transfer to the illuminator then view it (UVB) and photograph.

POST PCR PURIFICATION PROTOCOL

After every PCR the products remaining after running AGE that were to be applied to further detection techniques were purified using Performa® DTR Gel Filtration Cartridges (Edge Biosystems). They are 0.8 mL spin columns packed with a gel matrix optimized to effectively

remove dye dNTPs and low molecular weight materials from products. They also effectively remove DNA primers, fragments up 15 bases, buffers, biotin labeled nucleotides, isotope, dye terminators, and other assorted marker making them a suitable post-sequencing purification method as well. The protocol is as follows:

Performa® DTR Gel Filtration Cartridges (Edge Biosystems)

- Spin/centrifuge column at 3400 rpm for 2 min.
- Put column in fresh sterile tubes.
- Add samples (25-75 μL) to column.
- Centrifuge again at 3400 rpm for 2-4min (add ~ 10 μL of ddH20 and spin for 4 min. for optimal results.)

ABI BIG DYE TERMINATOR SEQUENCING PROTOCOL

The Sanger Dideoxy Method of DNA sequencing using ABI Big Dye Terminator v.3.1 Cycle Sequencing Kit protocols and the ABI 3100™ gel capillary electrophoresis in conjunction with Genescan™ software was employed as further confirmation that those fragments that were amplified were the desired product. The capillary electrophoresis was conducted by the Paleo-DNA Laboratory in Thunder Bay, Ontario and resulting electropherograms were analyzed using Bioedit 7.0.5.2. The sequencing also served as a tool to search the amplified products for any polymorphisms. The sequencing preparation was performed on ice in the dark. All reagents and samples kept on ice and vortexed and zipped before using. The sequencing protocol was performed each time as follows (Table 17):

Sanger Dideoxy Big Dye Terminator Sequencing Protocol

- Label a sterile 0.5 mL PCR tubes.
- Centrifuge post PCR purified extractions for 1 min. at 13000 rpm.
- Prepare the following PCR for each sample on ice (Table 17).

Table 17: BDT Sequencing Reagent volumes

Reagent	Volume
BDT Ready Rxn Mix	3 μL (or 4 μL)
Primer (forward or reverse) 10 μM	0.37 μL
Sample (purified)	According to gel band intensity (6-7 μL)
ddH_20	Enough to make a 10 μL reaction

- Prepare a master mix if sequencing 3 or more samples with the same primer.
- Vortex PCR tubes and briefly centrifuge.
- Program thermocycler: For 25-45 cycles.
    - Denaturation = 94 ^{0}C – 30 sec.
    - Annealing = 50 ^{0}C – 15 sec.
    - Extension = 72 ^{0}C – 4 min.
    - Hold at 4 ^{0}C.
- Purify using Performa® DTR Gel Filtration Cartridges.
- Desiccate completely at 30 °C.

ABI Prism® SNaPshot™ Primer Extension SNP Detection

The ABI Prism ABI Prism® SNaPshot™ primer extension method of SNP detection was used as the method to detect SNPs. All steps are prepared on ice and in the dark. The protocol was performed as follows:

ABI Prism® SNaPshot™ Primer Extension Protocol

- In order to prepare samples for the reaction an additional purification is requied. To 15 μL of post PCR purified template add 5 μL of SAP and 2 μl of Exo I and incubate at 37 °C for 1 hour followed by a 75 °C incubation for 15 minutes. Hold at 4 °C. These products are used for the SNaPshot™.
- SNaPshot protocol is performed on ice in the dark as follows (Table 18):

Table 18: SNaPshot™ Reagent volumes

SNaPshot™ ready reaction mix	5 μL
PCR product	3 μL
ddH20	1 μL
Primer(s) 10 μM	1 μL total

- The thermocycling parameters are as follows (Table 19):

Table 19: Thermocycler Parameters for the SNaPshot Reaction

Temperature (°C)	Time (s)
96 °C	10 s
50 °C	5 s
60 °C	30 s
4 °C	hold
cycles	25

- The reaction is ready to be detected following the addition of 1 uL of SAP incubated at 37 °C for 1 hour and 75 °C for 15 minutes, held at 4 °C as a post purification step.

DESIGN AND OPTIMIZATION OF A MULTIPLEX PRIMER EXTENSION SNP DETECTION SYSTEM

The ABI Prism® SNaPshot™ Primer Extension methodology was used to design single base extension primers in order to detect the presence of SNPs within the β-globin gene. It was chosen as it has multiplex capacity to increase throughput and decrease cost, it differentiates between products based on size and fluorescence, and it is compatible with already available capillary DNA sequencers (Syvanen, 2001). The compatibility is of importance as most assays, this one included, rely on detection methods that require highly specialized instrumentation, which are highly priced and not many laboratories cannot afford multiple systems (Kwok, 2001). This methodology uses the same gel capillary electrophoresis equipment and software as described in the Sanger Sequencing protocol described above.

Primers were designed to anneal to the already amplified, non-sequenced, and purified Amplicon III. The first primer HBBSNP6-1R is a reverse primer that will detect the first nucleotide at codon 6. It is at this position that the wild type genotype, or Hb Machida, or Hb C disease genotypes (homozygous or heterozygous) can be detected depending on which fluorescently labeled ddNTP is incorporated compliment to the genome. This primer produces a single base reverse complement extension product of 30 bp, where the incorporation of a C

(black) codes for wild type, the incorporation of a G (blue) codes for the presence of Hb Machida, and the incorporation of G (red) codes for Hb C.

The second primer HBBSNP6-2R is a reverse primer that will detect the second nucleotide at codon 6. It is at this position that the wild type, or Hb S, or Hb G Makassar, or Hb G San Jose disease genotypes (homozygous or heterozygous) can be detected depending on which fluorescently labeled ddNTP is incorporated compliment to the genome. The primer produces a single base reverse compliment extension product of 27 bp, where the incorporation of a T (red) codes for wild type, the incorporation of an A (green) codes for Hb S, the incorporation of a G (blue) codes for Hb G Makassar, and the incorporation of a C (black) codes for Hb G San Jose.

The third primer HBBSNP26 is a forward primer that will detect the first nucleotide at codon 26. It is at this position that the wild type or Hb E disease genotype (homozygous or heterozygous) can be detected depending on which fluorescently labeled ddNTP is incorporated identical to the genome. The primer produces a single base extension product identical to the genome of 24 bp, where the incorporation of a G (blue) codes for wild type, and the incorporation of an A (green) codes for Hb E.

The primers were optimized resulting in their addition in equimolar amounts totaling 1 μL of primer total. Once the reagents were titrated the pre-purification, the primer extension, and post purification methods were carried out as previously described in Protocols.

The specificity of the methodology increases as it progresses from step to step as the products are interrogated, annealed to, and amplified by a new set of oligonucleotide primers of different sequence that must exhibit complete homology with the products nucleotide sequence present. The sensitivity increases as the methodology progresses there are two amplification procedures whereby the second amplifies the initial amplification's products. The subsequent

amplification allows the detection and sequencing of a 1:1 000 000 diluted template. This is an increase of 100 fold detection capabilities than with a single amplification procedure. A loss of sensitivity is to be expected with each purification step as the products from the preceding step are diluted and a 100% efficiency of column purification is not guaranteed. Furthermore efficiency may decrease with each step as carry over reagents unused from the previous step may alter the chemical kinetics and product is lost. However these losses of sensitivity and efficiency are negligible if procedures are performed and the reagents and template are stored properly under the proper environmental conditions. When applying this methodology to degraded DNA samples contamination is of the utmost concern. Thus the precautions outlined in the Introduction and sterile working environments must be obeyed

This methodology was designed to amplify and analyze the first two exons of the β-globin gene in order to detect associated haemoglobinopathies. This family of gene and diseases was chosen as they obeyed the Mendelian pattern of inheritance, many of the diseases are prevalent and severe, the diseases effect global populations and their distribution, and the majority of diseases are attributed to SNPs. That being said this methodology can be modified. The cost, product loss, and risk of contamination can be decreased by removing the electrophoresis steps at the expense of losing a visual confirmation of desired amplification. However if well versed and confident in performing the methodology with accuracy and precision one can eliminate these electrophoresis steps and simply rely on sequencing to confirm specificity of amplification.

FINAL METHODOLOGICAL APPROACH

The following is a diagrammatic representation of the final developed methodology (Figure 7):

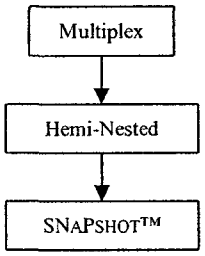


Figure 7: The Established and Optimized Methodological Approach

After all annealing temperatures, reagent concentrations, sensitivities were determined the optimized final methodology to be applied to modern and degraded medical and archaeological samples consists of a multiplex PCR, followed by a subsequent Hemi-nested PCR amplification, and finally by SNaPshot^{Tm} single based extension to be detected by capillary gel electrophoresis whereby the purified of products of the preceding procedure serve as the template for the subsequent procedure. Note that this methodology is to be applied to purified extracts. The optimal extraction/purification methodology is sample dependent.

CHAPTER 6: RESULTS

MULTIPLEX PCR

The Amplicon I primer set successfully and specifically amplified its desired amplicon of 256 bp. The annealing temperature was optimized to 63.0 °C (see Figure 8, lane 10) and the reagents were optimally titrated to an efficiency that successfully and specifically amplified a purified extract dilution of 1:10 000 (see Figure 9, lane 5). The following are the gel captures confirming the annealing temperature optimization and the dilution series respectively (Figures 8 and 9):

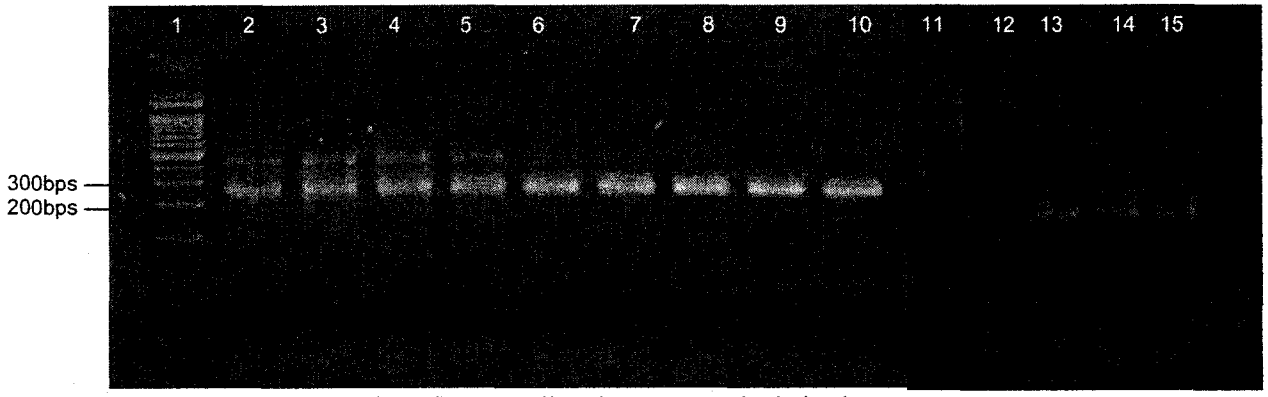


Figure 8: Amplicon I Primer Set Annealing Temperature Optimization
The purified extract used as template DNA was the sample extract as all subsequent reactions. The reagent concentrations were the same as those described in Table 2. The thermocycling parametres were the same as Table 5 with the exception of annealing temperatures which varied as follows: Lane 2, 55.0 °C; lane 3, 55.2 °C; lane 4, 55.7 °C; lane 5, 56.6 °C; lane 6, 57.8 °C; lane 7, 59.1 °C; lane 8, 60.5 °C; lane 9, 61.9 °C; lane 10, 63.1 °C; lane 13, 63.7 °C; lane 14, 64.0 °C; and lane 15, 65.5 °C. Lanes 1 and 11 are 1500 bp molecular size standard and lane 12 is a blank lane. Lane 10, approximated to 63.0 °C was determined visually to be the optimal annealing temperature as is had the best intensity with the least amount of non-specific product.

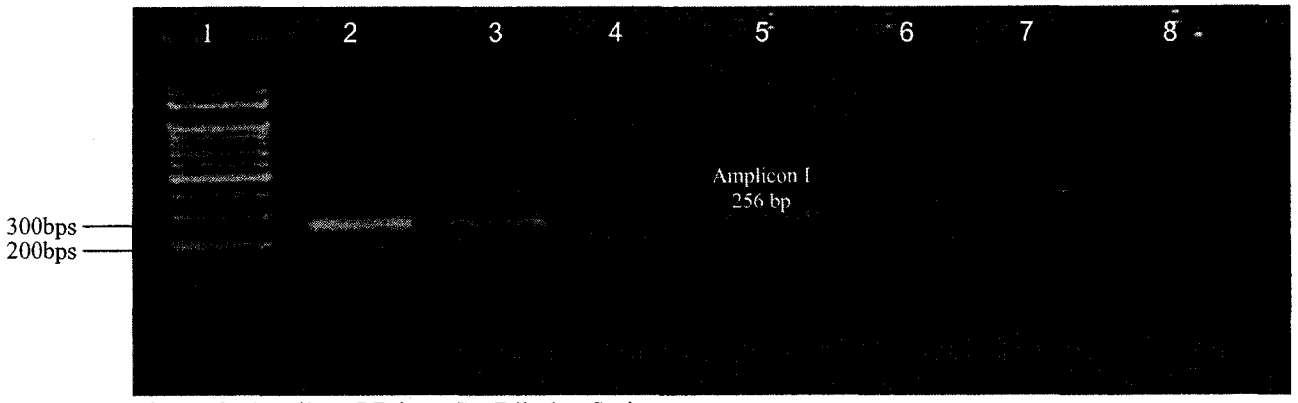


Figure 9: Amplicon I Primer Set Dilution Series
The purified extract used as template DNA was the sample extract as all previous and subsequent reactions. The optimized reagent concentrations and thermocycling parameters were the same as those described in Tables 4 and 5. Lane 1 contained the 1500 bp molecular size standard. The dilution series products were contained in the remaining lanes as follows: Lane 2, 1:10; lane 3, 1:100; lane 4, 1:1 000; lane 5, 1:10 000; lane 6, 1:100 000; lane 7, dilution negative; and lane 8, PCR negative.

The Amplicon II primer set successfully and specifically amplified its desired amplicon of 339bp. The annealing temperature was optimized to 65.0°C and the reagents were optimally titrated to an efficiency that successfully and specifically amplified a purified extract dilution of 1:1000. The following is the gel capture confirming the annealing temperature optimization and the dilution series (Figure 10):

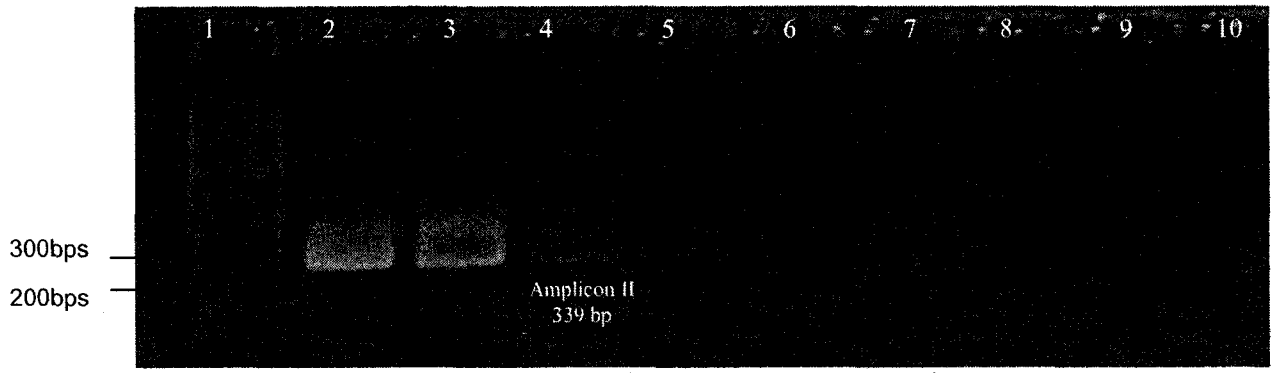


Figure 10: Amplicon II Primer Set Optimized Annealing Temperature Dilution Series
The purified extract used as template DNA was the sample extract as all previous and subsequent reactions. The optimized reagent concentrations and thermocycling parameters were the same as those described in Tables 6 and 7. Lane 1 contained the 1500 bp molecular size standard. The dilution series products were contained in the remaining lanes as follows: Lane 2, 1:10; lane 3, 1:100; lane 4, 1:1 000; lane 5, 1:10 000; lane 6, 1:100 000; lane 7, 1:1 000 000; lane 8, blank; lane 9, PCR negative; and lane 10, dilution negative.

The amplicons in figure 10 are significant smearing and the migration of the size standard does not match entirely for its desired length, nor do the desired amplicons in figures 8 and 9. This smearing is attributed to too much DNA, which will not be an issue when applied to degraded samples. The undesirable length can be attributed to expired and degraded molecular size standard causing a difference in migration. A difference in migration can even be appreciated between samples as AGE is not a uniformly synthetic migration media and the comb that produces the lanes may not have been truly straight. The difference is millimetres. Though the visual confirmation is not optimal the products have been confirmed as specific through sequencing (Appendix C:B).

As previously stated the sequencing of optimized amplifications and dilution series were conducted to ensure specificity of amplifications. Though a discrepancy was observed on AGE the fact that the desired sequences were obtained one can attribute said gel observations to the issues discussed above. The optimized Amplicon I amplification was sequenced despite being of desired length, as was its dilution series (Appendix C:A). The sequence matched 218 out of 219 nucleotides with 100% homology to the reference sequence. This corresponds to 99.54% homology. The discrepancy occurs at nucleotide position 1620, whereby a C → T heterozygous state exits at the third amino acid in exon 1's codon which results in a silent mutation. This extract's heterozygous state is confirmed by the reverse sequence and subsequent dilution sequences. The 1:1000 purified dilution series product, when sequenced, confirmed the previous results and displayed the same homology.

The optimized Amplicon II was sequenced, as was its dilution series to ensure the desired amplicon was amplified. There was 100% homology when compared to the reference sequence (Appendix C:B). The forward primer was not capable of sequencing the entire exon however the

reverse did for the reasons discussed in Protocols: The Design and Optimization of Multiplex PCR (Appendix C, B). Note that the purity of the negatives in both primer sets did not produce any discernable sequence.

The multiplex PCR was optimized and its sensitivity was determined upon confirmation that the singleplex reactions of the Amplicon I and II primer sets were specific and they could detect dilutions of 1:10 000 (Figure 9, lane 5) and 1:1 000 (Figure 10, lane 3) respectively. Its optimal annealing temperature was determined to be 63.0°C (Figure 11, lane 4) and was sensitive enough to detect a dilution of 1:10 000. The following is a gel capture of the multiplexed optimized annealing temperature (Figure 11):

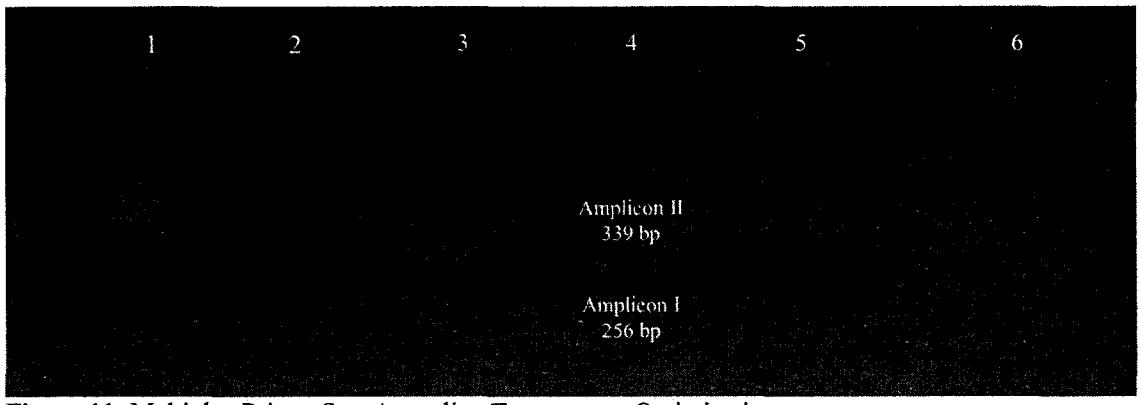


Figure 11: Multiplex Primer Sets Annealing Temperature Optimization
The purified extract used as template DNA was the sample extract as all subsequent reactions. The reagent concentrations were the same as those described in Table 2. The thermocycling parametres were the same as Table 5 with the exception of annealing temperatures which varied as follows: Lane 2, 57.8 °C; lane 3, 60.5 °C; lane 4, 63.0 °C; and lane 5, 65.0 °C. Lane 1 contains 1500 bp molecular size standard and lane 6 contains a PCR negative. Lane 4, 63.0 °C was determined visually to be the optimal annealing temperature as is had the best intensity with the least amount of non-specific product.

The gel capture (Figure 11) is of poor resolution. The molecular size standard (see Figure 11, lane 1) is once again of poor resolution. In addition the AGE did not permit the size standard and PCR products to migrate at the desired rate. However two distinct bands are visible and rather than confirming the correct amplification visually it was confirmed through sequencing

(Appendix C:A & B). The following is a gel capture of the multiplexed dilution series (Table 12):

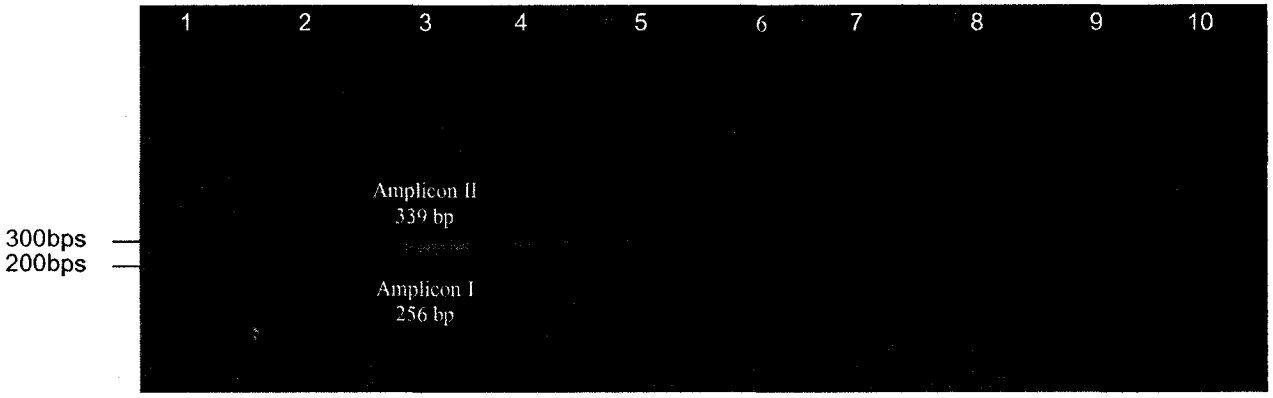


Figure 12: Multiplex Primers Sets Dilution Series
The purified extract used as template DNA was the sample extract as all previous and subsequent reactions. The optimized reagent concentrations and thermocycling parameters were the same as those described in Tables 8 and 9. Lane 1 contained the 1500 bp molecular size standard. The dilution series products were contained in the remaining lanes as follows: Lane 3, 1:10; lane 4, 1:100; lane 5, 1:1 000; lane 6, 1:100 000; lane 7, 1:10 000; and lane 8, 1:1 000 000. Lane 2 was blank and did not contain any sample. Lane 9 contained the dilution negative and lane 10 contained the PCR negative.

Once again the amplicons in Figure 12 are difficult to distinguish between each other. This can be attributed to the low resolution capabilities of the AGE and migratory behavioural differences between the size standard and the products. The lanes in which the templates were added are not level with each other creating a difference in their vertical displacement.

The sequences performed on the optimized amplification and dilution series products as low as 1:10 000 produced the same results as its singleplexed counterparts thus confirming specific amplification despite visual discrepancies observed on the AGE (Appendix C: A & B). This designed and optimized multiplexed methodology using Amplicon I and II primer sets was capable of successfully and specifically amplifying the desired first two β-globin gene exons and thus any inherited genetic variant with a level of sensitivity capable of detecting DNA dilutions

of 1:10 000 (Figure 12, lane 7). All negatives, both dilution and PCR negatives (Figure 12, lanes 9 and 10) were assured of purity as no discernable sequences could be produced.

HEMI-NESTED MULTIPLEX PCR

The Amplicon III primer set successfully and specifically amplified its desired amplicon of 227bp (Figure 13, lane 6). The annealing temperature was optimized to 60.0°C (Figure 13, lane 8) and the reagents were optimally titrated to an efficiency that successfully and specifically amplified a purified extract dilution of 1:100 000 (Figure 14, lane 4). The following are the gel captures confirming the annealing temperature optimization and the dilution series (Figure 13):

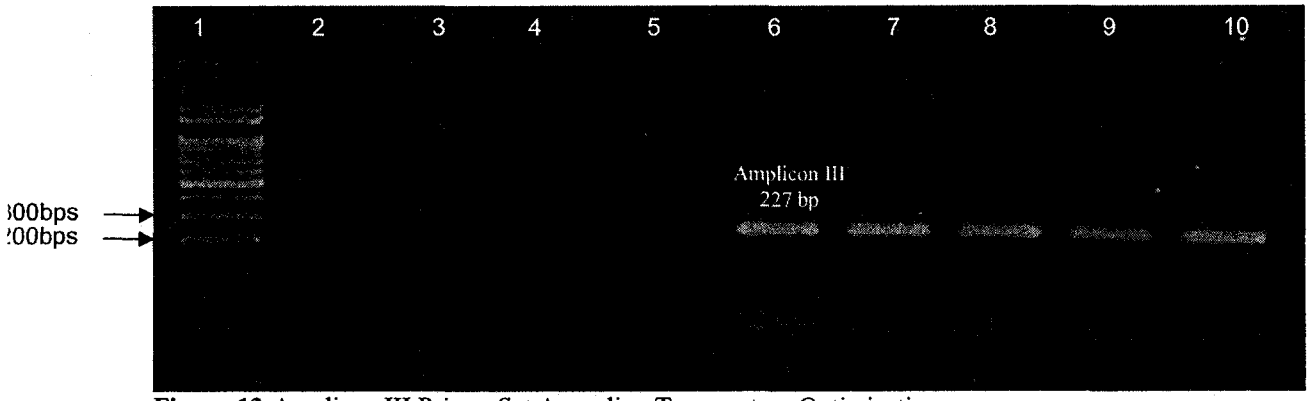


Figure 13: Amplicon III Primer Set Annealing Temperature Optimization:
The purified extract used as template DNA was the sample extract as all subsequent reactions. The reagent concentrations were the same as those described in Table 2. The thermocycling parametres were the same as Table 5 with the exception of annealing temperatures which varied as follows: Lane 6, 55.0 °C; lane 7, 57.8 °C; lane 8, 60.5 °C; lane 9, 63.0 °C; and lane 10, 65.0 °C. Lane 1 contained a 1500 bp molecular size standard and lane 5 contains a PCR negative. Lane 2 is blank and lanes 3 and 4 were samples not from this study. The annealing temperature, based on the results observed visually in lane 8 to be approximated to 60.0 °C as it had the best intensity with the least amount of non-specific product.

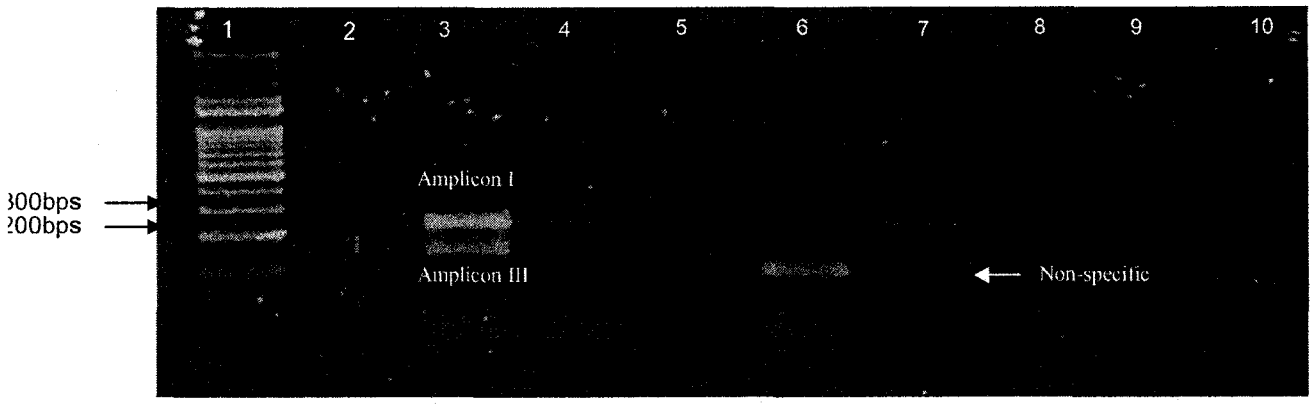


Figure 14: Amplicon III Primer Set Dilution Series
The purified extracts used as template DNA were the corresponding dilution products from Figure 9. The optimized reagent concentrations and thermocycling parameters were the same as those described in Tables 10 and 11. Lane 1 contained the 1500 bp molecular size standard. The dilution series products were contained in the remaining lanes as follows: Lane 3, 1:10 000; and lane 4, 1:100 000. Lanes 2 and 5 were blank. Lane 6 contained the purification negative, which is the PCR negative product from lane 8, Figure 9 but added to the same PCR parameters as those described in Tables 10 and 11. Lane 7 contained the PCR negative and the remaining lanes are blank.

Both Amplicon III and Amplicon I can be observed in the dilution series gel capture (Figure 14, lanes 3 and 4). This is to be expected as the amplified template from Amplicon I primer sets is at far greater concentration than a normal initial PCR template and thus not enough of the hemi-nested primer was added. Though if more primer was to be added it would change the concentrations of the other PCR reagents and the reaction would no longer be optimized. Non-specific amplicons are observed below the desire product in Figure 14, lane 4 as well as in 6 and 7 however it is attributed to primer extension products. The negatives produced non specific products however when sequenced no discernable sequence could be identified.

The Amplicon IV primer set successfully and specifically amplified its desired amplicon of 234bp and the migration of the size standard matched that of the desired products. This reaction was a success as the specificity was confirmed by sequencing the products. The annealing temperature was optimized to 65.0°C (Figure 15, lane 6) and the reagents were optimally titrated to an efficiency that successfully and specifically amplified a purified extract dilution of 1:1 000

(Figure 16, lane 5). The following are the gel captures confirming the annealing temperature optimization and the dilution series respectively (Figures 15 and 16):

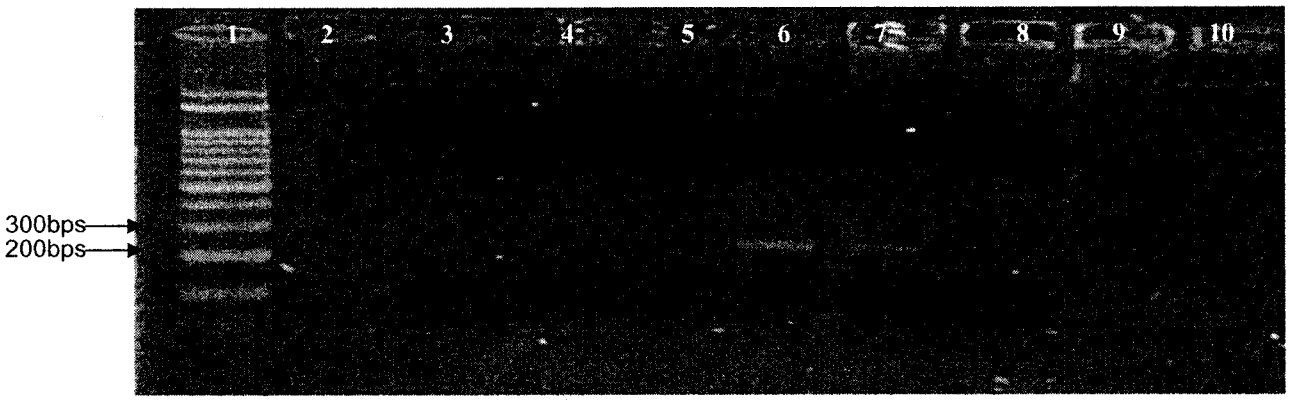


Figure 15: Amplicon IV Primer Set Annealing Temperature Optimization
The purified extract used as template DNA was the sample extract as all subsequent reactions. The reagent concentrations were the same as those described in Table 2. The thermocycling parametres were the same as Table 5 with the exception of annealing temperatures which varied as follows: Lane 3, 57.8 °C; lane 4, 60.1 °C; lane 5, 63.2 °C; lane 6, 65.1 °C; and lane 7, 66.4 °C. Lane 1 contained a 1500 bp molecular size standard and lanes 2, 9, and 10 are blank lanes. Lane 8 contained the PCR negative. Lane 6, approximated to 65.0 °C was determined visually to be the optimal annealing temperature as is had the best intensity with the least amount of non-specific product.

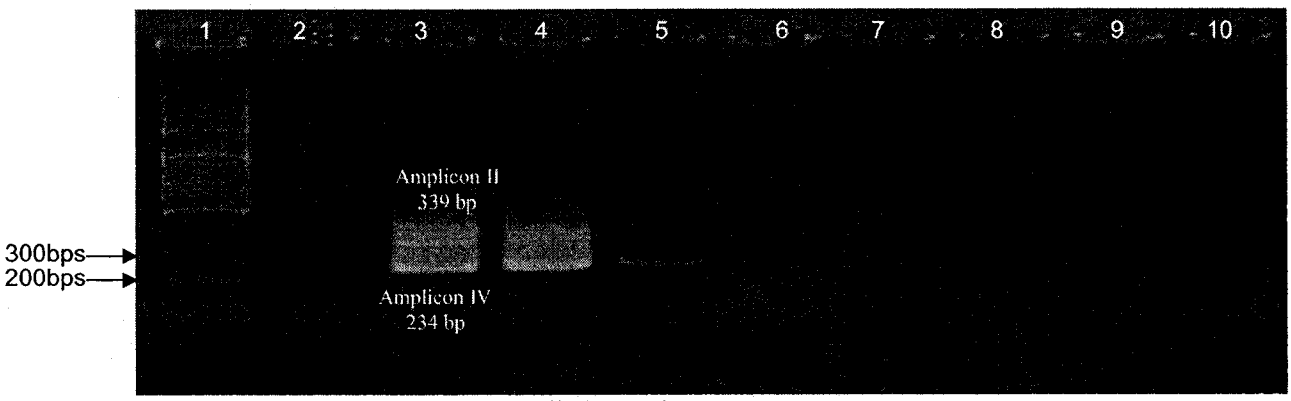


Figure 16: Amplicon IV Primer Set Dilution Series
The purified extracts used as template DNA were the corresponding dilution products from Figure 10. The optimized reagent concentrations and thermocycling parameters were the same as those described in Tables 12 and 13. Lane 1 contained the 1500 bp molecular size standard. The dilution series products were contained in the remaining lanes as follows: Lane 3, 1:10; and lane 4, 1:100; lane 5, 1:1 000; lane 6, 1: 100 000; lane 7, 1:10 000 Lanes 2 and 8 were blank. Lane 9 contained the purification negative, which is the PCR negative product from lane 9, Figure 10 but added to the same PCR parameters as those described in Tables 10 and 11. Lane 10 contained the PCR negative and the remaining lanes are blank.

The desired sized amplicon of 234bp was produced and though there is significant smearing the products have been confirmed as specific through sequencing. This smearing is attributed to too much DNA carried over from the products of Amplicon II dilution series. This is not of concern as these methodologies are designed to work with degraded DNA and would produce products more similar to those in Figure 16, lane 5. Both Amplicon IV and Amplicon II can be observed in the dilution series gel capture, as a band of approximate length of 339 bp is observed in Figure 16, lanes 3 and 4. This is to be expected as primers from the initial amplification may still be present or not enough of the hemi-nested primer was added. It did successfully detect a dilution of 1:10 000 (Figure 16, lane 7)

Thus the optimized amplifications and dilution series were sequenced (Appendix C:A & B) and they confirmed the sequences obtained by the dilution multiplex PCR (Figure 12). The primer sets for Amplicons III and IV successfully produced discernable sequences (Appendix C:A & B) for dilutions of 1:100 000 (Figure 14, lane 4) and 1:10 000 (Figure 16, lane 7) respectively. No discernable sequences were produced from the negatives.

The hemi-nested multiplex PCR was optimized and its sensitivity was determined upon confirmation that the singleplex reactions of the Amplicons III and IV primer sets were specific and they could detect dilutions of 1:100 000 (Figure 14, lane 4) and 1:10 000 (Figure 16, lane 3) respectively. Its optimal annealing temperature was determined to be 60.0 °C (Figure 17, lane 5) and was sensitive enough to detect a dilution of 1:1 000 000 (Figure 18, lane 9). The following is a gel capture of the hemi-nested multiplexed optimized annealing temperature (Figure 17):

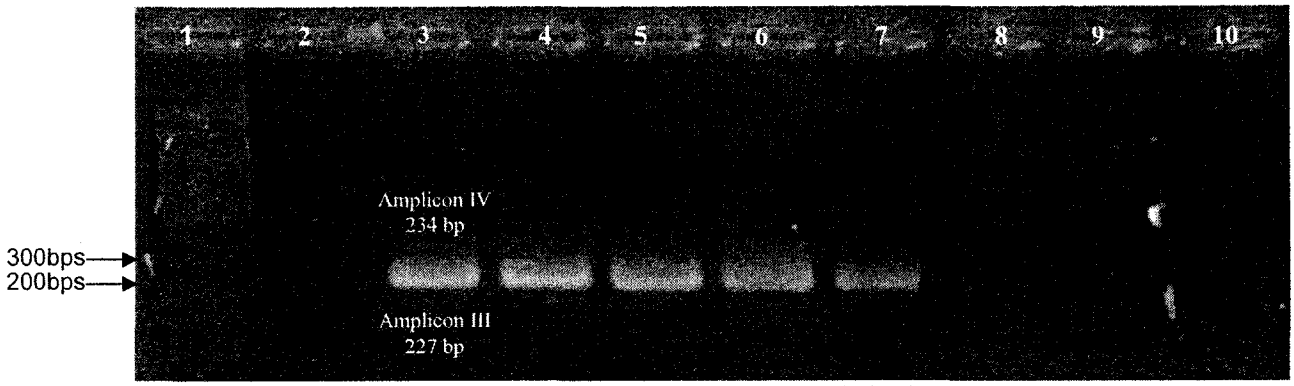


Figure 17: Hemi-Nested Primer Sets Annealing Temperature Optimization
The purified extract used as template DNA was the sample extract as all subsequent reactions. The reagent concentrations were the same as those described in Table 2. The thermocycling parametres were the same as Table 5 with the exception of annealing temperatures which varied as follows: Lane 3, 55.0 °C; lane 4, 57.8 °C; lane 5, 60.5 °C; lane 6, 63.0 °C; and lane 7, 65.0 °C. Lane 1 contained a 1500 bp molecular size standard and lanes 2, 8, and 10 are blank lanes. Lane 9 contained the PCR negative. Lane 5, approximated to 60.0 °C was determined visually to be the optimal annealing temperature as is had the best intensity with the least amount of non-specific product.

Though the resolution of the size standard is poor the amplicons are of great enough thickness and fluorescent intensity to assume that both amplifications products, 227 and 234 bp, are present. The resolution does not permit differentiation between two amplicons separated in length by only seven base pairs. This was distinguished from two much DNA by sequencing both products (Appendix C:A & B). Though an artifact is present near lane 9, Figure 17, it is clear to see that the PCR negative has not been contaminated.

The purified multiplexed dilution products from Figure 12 were used as templates for the hemi-nested multiplex reaction in order to observe an increase in sensitivity capabilities. The PCR negative from the multiplex dilution series (Figure 12, lane 10) was again purified and reagents added in order to serve as a purification negative to track contaminated reagents or improper handling techniques (Figure 18, lane 9). In addition a positive undiluted sample was added (Figure 18, lane 1) and as always the PCR negative (Figure 18, lane 10). The following is the gel capture of the hemi-nested multiplex applied to purified multiplex dilution products (Figure 18):

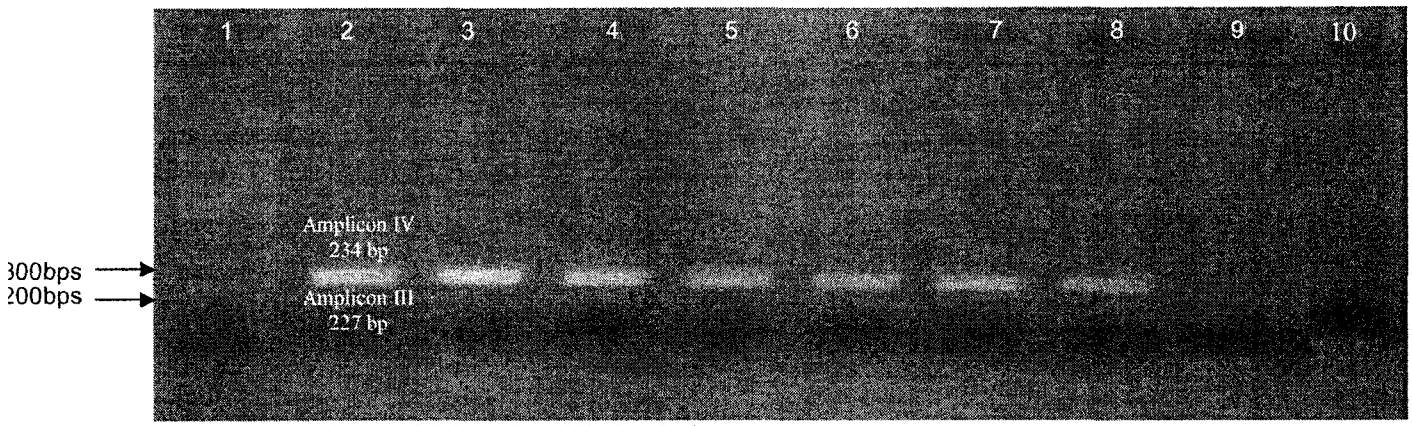


Figure 18: Hemi-Nested Primer Sets Dilution Series
The purified extracts used as template DNA were the corresponding dilution products from Figure 12.The optimized reagent concentrations and thermocycling parameters were the same as those described in Tables 14 and 15. Lane 1 contained the 1500 bp molecular size standard. The dilution series products were contained in the remaining lanes as follows: Lane 3, 1:10; lane 4, 1:100; lane 5, 1:1 000; lane 6, 1:10 000; lane 7, 1:100 000; lane 8, 1:1 000 000. Lane 1 contained the same undiluted extract used in all optimization and annealing temperature reactions. Lane 9 contained the purification negative, which is the PCR negative product from lane 10, Figure 12 but added to the same PCR parameters as those described in Tables 14 and 15. Lane 10 contained the PCR negative.

There are clearly two amplification products of the same nucleotide length present. The negatives were not contaminated and the desired amplicon lengths of 227 and 234 bp were amplified (Figure 18, lane 2). Note that the PCR negative in lane 10 (Figure 18) was ran on a separate gel and thus a migration difference exists between the running dye when compare to the first nine lanes. Sequencing confirmed these results with those previously conducted in both this step and step I.

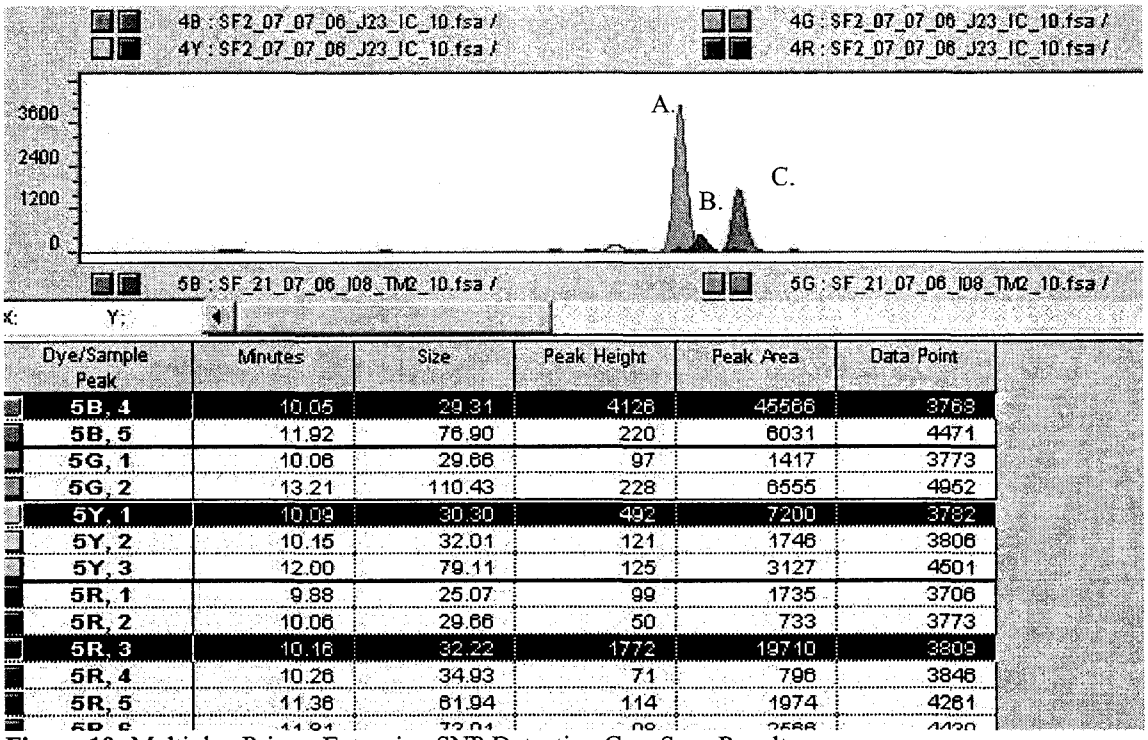
Overall the hemi-nested successfully improved sensitivity. The subsequent amplification allowed the singleplex primer sets to detect and sequence a 10 fold decrease in template concentration. The subsequent hemi-nested multiplex amplification produced a detectable amplicon from a template dilution of 1:1 000 000 (Figure 18, lane 8), that was of similar intensity as the undiluted product (Figure 18, lane 1). This represents a 100 fold increase in sensitivity compared to the detectable products produced my multiplex alone (Figure 12). The sensitivity of the hemi-nested methodology has naturally been increased as each amplicon has to

specifically hybridize to three primers instead of just two. Thus the sensitivity and specificity have been increased with the design and optimization of the hemi-nested multiplex methodology to follow an initial multiplex PCR.

MULTIPLEX PRIMER EXTENSION SNP DETECTION SYSTEM

The primers were optimized resulting in there addition in equimolar amounts totaling 1 μL of primer total. Once the reagents were titrated the pre-purification, the primer extension, and post purification methods were carried out as described previously in Protocols.

The methodology was successfully optimized as a multiplex. The results were interpreted by Genescan™. It produced homozygous single base extension products corresponding to wild type genotype for all positions. The peaks however, due to the chemical dynamics and preferential annealing, could not be optimized to provide equal detection intensities with this sample. The fragments produced unfortunately do not correspond with their expected mobility. HBBSNP26 is detected at a length corresponding to 29.31 bp instead of 24 bp. HBBSNP6-1R is detected accurately, and HBBSNP6-2R is detected at a length corresponding to 32.22 bp. This however is not of concern, and these results are considered accurate as there is a mobility error of ± 3-6bp and they are within their range of error. The following is the results obtained on GeneScan post-primer extension multiplex optimization (Figure 19):



Dye/Sample Peak	Minutes	Size	Peak Height	Peak Area	Data Point
5B, 4	10.05	29.31	4128	45566	3768
5B, 5	11.92	76.90	220	6031	4471
5G, 1	10.06	29.66	97	1417	3773
5G, 2	13.21	110.43	228	6555	4952
5Y, 1	10.09	30.30	492	7200	3782
5Y, 2	10.15	32.01	121	1746	3806
5Y, 3	12.00	79.11	125	3127	4501
5R, 1	9.88	25.07	99	1735	3706
5R, 2	10.06	29.66	50	733	3773
5R, 3	10.16	32.22	1772	19710	3809
5R, 4	10.26	34.93	71	796	3846
5R, 5	11.36	61.94	114	1974	4261

Figure 19: Multiplex Primer Extension SNP Detection GeneScan Results
A. represents HBBSNP26 single base extension product coding for wild type at its corresponding nucleotide position, B. represents HBBSNP6-1R single base extension product coding for the reverse compliment of wild type at its corresponding nucleotide position, and C. represents HBBSNP6-2R single base extension product coding for the reverse compliment of wild type at its corresponding nucleotide position.

This design and optimization was a success. This methodology once again increases specificity as it incorporates and requires yet another new primer per SNP position to be hybridized. Any given SNP at this point has been interrogated by four different primers, one of which. HBBIIAR anneals twice. The sensitivity in terms of lower concentration of template would not be expected to be increased as an additional amplification has not been introduced. Though the detection sensitivity may be decreased with the addition of more primers as there is a finite amount, the necessary fragment size to anneal and detect has been reduced to the size of the designed primer plus one nucleotide. Therefore the potential to detect smaller, more degraded fragments exists and thus may prove to increase the sensitivity by detecting previously undetectable samples.

CHAPTER 7: DISCUSSION

It is through this project that the objective to design and optimize a multi-step methodology of increasing specificity and sensitivity on modern samples that can be applied to medically archived, ancient, and degraded samples was accomplished. The designed methodology to be applied is a three step process: 1) Multiplex PCR, 2) Hemi-nested PCR, and 3) Single base extension detection of SNPs using the ABI Prism® SNaPshot™ Primer Extension kit. The specificity was increased as predicted by annealing multiple primers sets to detect single gene sequences and the sensitivity was increased by applying amplified products to a second amplification step. This increase in sensitivity and specificity along with the SNaPshot™ protocol is hypothesized to allow this methodology to detect the presence of genetic mutations with an emphasis on SNP detection.

The specificity of the methodology increases as it progresses from step to step as the products are interrogated, annealed to, and amplified by a new set of oligonucleotide primers of different sequence that must exhibit complete homology with the products nucleotide sequence present (Dieffenbach et al., 1993). The sensitivity increases as the methodology progresses there are two amplification procedures whereby the second amplifies the initial amplification's products (McPherson, 2000; Mullis et al., 1986; Ruano et al., 1989). This results in the detection and sequencing of template in a 1:1000000 dilution, an increase of 100 fold detection capabilities than with a single amplification procedure. A loss of sensitivity is to be expected with each purification step as the products from the preceding step are diluted and a 100% efficiency of column purification is not guaranteed (Stewart et al., 2003). Furthermore efficiency may decrease with each step as carry over reagents unused from the previous step may alter the chemical kinetics and product is lost (McPherson, 2000). However these losses of sensitivity

and efficiency are negligible if procedures are performed and the reagents and template are stored properly under the proper environmental conditions. When applying this methodology to degraded DNA samples contamination is of the utmost concern (Hofreiter et al., 2001; Willerslev and Cooper, 2005). Thus the precautions outlined in Chapter 2 and sterile working environments must be obeyed.

Design and optimization of primer sets was of paramount importance. Though described more thoroughly in Chapter 3 a balance must be established between specificity and efficiency (Dieffenbach et al., 1993). These primer sets were designed for this methodology to amplify and analyze the first two exons of the β-globin gene in order to detect associated haemoglobinopathies. This family of gene and diseases was chosen as they obeyed the Mendelian pattern of inheritance, many of the diseases are prevalent and severe, the diseases effect global populations and their distribution, and the majority of diseases are attributed to SNPs (Collins, 1984; Weatherall and Clegg, 2001; WHO, 1989). That being said this methodology can be modified. The cost, product loss, and risk of contamination can be decreased by removing the electrophoresis steps at the expense of losing a visual confirmation of desired amplification. However if well versed and confident in performing the methodology with accuracy and precision one can eliminate these electrophoresis steps and simply rely on sequencing to confirm specificity of amplification. Furthermore if one is only concerned with, at least initially, detecting the presence of the most prevalent and severe β-globinopathies with in a specimen then the sequencing of multiplex and hemi-nested PCR products can be omitted, and one can rely solely the detection capabilities associated with the designed ABI Prism® SNaPshot™ Primer Extension methodology. Considerable time and cost can be avoided.

This methodology can be applied to detect any known small scale genetic mutations. Provided that the mutated nucleotide sequence is known, and primers are designed and optimized through application to this three step methodology, it is hypothesized that the sensitivity and specificity will be capable of detecting and obtaining genetic sequences from increasingly degraded samples. Furthermore by multiplexing the different primer sets one can expand upon this study to include multiple genetic mutations in an all inclusive study of choice. This naturally has medical and clinical implications. The multiplexed methodology reduces cost and time thus reducing the diagnostic time and increasing overall patient care. Archived and fragmented patient samples from previous tests can now be repeatedly tested with new disease specific primers reducing the impact on the patient and reducing potential travel expenses, a key issue in regional care centres when patients live in rural and remote communities.

The increase in specificity and sensitivity of this methodology now provides a tool to investigate degrade medically archived and archaeological samples. This methodology can easily be adapted in order to validate any historical diagnosis from medically archived samples, thus validating early medical studies. Furthermore it can now be applied to detect the biomarkers in ancient and degraded samples to validate any previous morphological tissue findings or attempt to establish the evolution of human diseases but their impact on cultures and populations (Faerman et al., 2000; Filon et al., 1995). Archaeologically one could apply multiple primer sets designed to detect multiple diseases in order to investigate the presence of complete spectrums of disease in ancient populations. The archaeological and clinical applications were designed to

The single based primer extension methodology has multiplex capabilites and through its automation has the potential to rapidly analyze multiple samples at multiple locations (Syvanen,

2001). Thus the SNP detection methodology can interrogate large modern amplified fragments and data can be applied to linkage mapping or association studies (Kwok, 2001). It is the presence of SNP patterns within non-coding regions that are thought to be in association with genes that contribute to the polygenic or multifactorial common diseases whose inheritance patterns have yet to be fully identified. These mapping studies can be utilized within the fields of population genetics and forensics. Many SNPs are conserved within certain populations and are stable enough biomarkers to be used to conduct population migration studies. As SNPs do occur at high frequencies and are thought to be the factors that contribute to our 0.1% genetic individuality (Li and Sadler, 1991; Wang et al., 1998; Wang and Todd, 2003). The sensitivity and specificity of this study allows its application to potentially degraded samples for forensic identification.

CHAPTER 8: CONCLUSIONS AND FUTURE DIRECTIONS

A multi-stepped, multiplexed methodology of increasing sensitivity and specificity was designed and optimized to detect genetic diseases with an emphasis towards SNPs. It consists of three steps: multiplex PCR of already extracted and purified DNA, hemi-nested PCR of purified multiplex products, and sequencing of amplifications, followed by single based extension primer detection applied to hemi-nested products (see Figure 7). The sensitivity increases with each subsequent amplification and the specificity increases as each subsequent step relies on 100% homology between the template sequences and the new primer sets. Hence the methodology can be applied to degraded and low copy DNA samples. The methodology was designed and modeled to detect the presence of genetic mutations within the first two exons of the β-globin gene. These small scale mutations referred to as haemoglobinopathies, served as the perfect mutations to design and optimize a methodology around. They follow Medelian patterns of inheritance, have simple molecular genetics, are prevalent, and those that are severe enough to affect global populations are the result of a SNP. All designed primer sets were optimized and when viewed by electrophoresis produced amplicons of desired length. The amplicon specificity was validated by sequencing and comparing sequences to the reference sequence.

It was determined that through subsequent amplifications and primer hybridizations that a methodology of increasing sensitivity and specificity was obtained that is capable of detecting genetic disease in degraded tissue samples. In doing so it is hope of the study that this methodology will be applied to ancient samples to expand upon previous studies and bridge the gap between the diagnosis and study of β-haemoglobin variants in both modern and ancient specimens (Faerman et al., 2000; Filon et al., 1995). In accomplishing so it is hypothesized that further contribute to the research they have started.

Further methodological validation, however, is required. This reproduction of this study must be completed. Furthermore it must be applied to a larger sample size of human DNA that individual's possessing different genotypic states in order determine the trial sensitivity and specificity of the techniques used. Once established the methodology should be applied to medically archived specimens to validate both procedures. Once conducted the methodology is to be applied to ancient and degraded samples from areas with high prevalence for haemoglobinopathies and/or exhibit pathophysiological and morphological features of haemolytic anemias. Finally this study should be further expanded to include diseases that exhibit similar pathophysiology findings, inheritance patterns, or are associated with similar geographical distribution such as tuberculosis, leprosy, osteoarthritis, or malaria thus creating a larger more inclusive diagnostic test to reduce time and cost of analysis (Gernaey et al., 2001; Hershkovitz et al., 1997; Papathanasiou, 2005).

APPENDIX A

Table 20: Optimized Proteinase K DNA Extraction (Applied Biosystems)

Reagents	Lot #
TNE	H17-06 11/03/06 TM
20% SDS	TM 06/01/06
0.39M DTT	TM 06/01/06
Proteinase K	43119

Table 21: E.Z.N.A Cycle Pure (United Bioformatica Inc.)

Reagents	Lot #
1.5mL tubes	5105
HiBind® DNA spin column	B6013
XP1	40505
SPW	032805

Table 22: Running/Loading Agarose Gel Electrophoresis

Reagents	Lot #
6X Loading dye	H13-06
100bp Ladder	A0126
EtBr	RL 03/11/05

Table 23: Performa® DTR Gel Filtration Cartridges (Edge Biosystems)

Reagents	Lot #
Kit	144062
Collection tubes	143766
Cartridges	143862

Table 24: PCR Reagents (Invitrogen)

Reagents	Lot#
PCR Buffer	1143243
dNTP	2-04
$MgCl_2$	1215526
Platinum® *Taq* DNA Polymerase	1233571
HBBIIF	6831
HBBIIAR	6009
HBBIIIF	6833
HBBIIIAR	6028
HBB1F	6609
HBBIIINR	6029
HBB3F	6611
HBBR	6791

APPENDIX B

Table 25: Sequence of Primers Used

Primer	Sequence 5' to 3'	Genome Position	Desired Amplicon bp	Experi-mental T_a (°C)	Source	NCBI Accession	Nucleotide Position
Detection PCR							
HBB3F	TCTTAGAGGAGGGCTGAGGGTTT	Partial β-Hb Exon1	315	63.0	Beaulne, 2004	M34058	1367
HBBR	ACTTCATCCACGTTCACCTTGCCCC	Partial β-Hb Exon1			Beaulne, 2004	M34058	1682
Amplicon I							
HBBIIF	GGCAGAGCCATCTATTGCTTAC	B-Hb Exon 1	256	63.0	This Study	M34058	1542
HBBIIAR	GTCATAGGCACTGACTCTCTCT	B-Hb Exon 1			This Study	M34058	1798
Amplicon II							
HBBIIIF	GACCCAGAGGTTCTTTGAGTC	B-HB Exon 2	339	65.0	This Study	M34058	1855
HBBIIIAR	CTGAGACTTCCACACTGATGC	B-HB Exon 2			This Study	M34058	2194
Amplicon III							
HBB1F	TCTGACACAACTGTGTTCACTAGCAAC C	B-Hb Exon 1	227	60.0	This Study	M34058	1571
HBBIIAR	GTCATAGGCACTGACTCTCTCT	B-Hb Exon 1			This Study	M34058	1798
Amplicon IV							
HBBIIIF	GACCCAGAGGTTCTTTGAGTC	B-HB Exon 2	236	65.0	This Study	M34058	1855
HBBIIINR	GGAAAGAAAACATCAAGGGTCCC	B-HB Exon 2			This Study	M34058	2089
SNP Single Base Primer Extension Priners							
Primer	Sequence 5' to 3'	Genome Postion	Desired Amplicon bp	Experi-mental T_a (°C)	Source	NCBI Accession	SNP Nucleotide Position
HBBSNP6-1R	CCCACAGGGCAGTAACGGCAGACTTCT CCT	B-Hb codon 6	30	60.0	This Study	M34058	1630
HBBSNP6-2R	CACAGGGCAGTAACGGCAGACTTCTCC	B-Hb codon 6	27	60.0	This Study	M34058	1631
HBBSNP26	GTGAACGTGGATGAAGTTGGTGGT	B-Hb codon 26	24	60.0	This Study	M34058	1751

APPENDIX C

The following are the sequences produced by primers used in this study and are compared to their corresponding reference sequence. Any discrepancies are highlighted in red.

A. Sequence Information: Starting position 1541 accession M34058

```
Exon1     GGGCAGAGCCATCTATTGCTTACATTTGCTTCTGACACAACTGTGTTCACTAGCAACCTCAA
HBBIIF    ----------------------------------------------------------ACTGTGTTCACTAGCAACCTCAA
HBB1F     ---------------------------------------------------------------------------------------------------------
HBBIIAR   ----------------------------------------------------TTCTGACACAACTGTGTTCACTAGCAACCTCAA

Exon1     ACAGACACCATGGTGCACCTGACTCCTGAGGAGAAGTCTGCCGTTACTGCCCTGTGGGGCA
HBBIIF    ACAGACACCATGGTGCANCTGACTCCTGAGGAGAAGTCTGCCGTTACTGCCCTGTGGGGCA
HBB1F     ACAGACACCATGGTGCANCTGACTCCTGAGGAGAAGTCTGCCGTTACTGCCCTGTGGGGCA
HBBIIARACAGACACCATGGTGCANCTGACTCCTGAGGAGAAGTCTGCCGTTACTGCCCTGTGGGGCA

Exon1     AGGTGAACGTGGATGAAGTTGGTGGTGAGGCCCTGGGCAGGTTGGTATCAAGGTTACAAGA
HBBIIF    AGGTGAACGTGGATGAAGTTGGTGGTGAGGCCCTGGGCAGGTTGGTATCAAGGTTACAAGA
HBB1F     AGGTGAACGTGGATGAAGTTGGTGGTGAGGCCCTGGGCAGGTTGGTATCAAGGTTACAAGA
HBBIIARAGGTGAACGTGGATGAAGTTGGTGGTGAGGCCCTGGGCAGGTTGGTATCAAGGTTACAAGA

Exon1     CAGGTTTAAGGAGACCAATAGAAACTGGGCATGTGGAGACAGAGAAGACTCTTGGGTTTCT
HBBIIF    CAGGTTTAAGGAGACCAATAGAAACTGGGCATGTGGAGACAGAGAAGACTCTTGGGTTTCT
HBB1F     CAGGTTTAAGGAGACCAATAGAAACTGGGCATGTGGAGACAGAGAAGACTCTTGGGTTTCT
HBBIIARCAGGTTTAAGGAGACCAATAGAAACTGGGCATGTGGAGACAGAGAAGACTCTT----------

Exon1     GATAGGCACTGACTCTCTCTGCCTATTGGTCTATTTTCCCACCCTTAGGCTGCTG----------
HBBIIF    GATAGGCACTGAC-------------------------------------------------------------------------------------
HBB1F     GATAGGCACTG---------------------------------------------------------------------------------------
HBBIIAR-------------------------------------------------------------------------------------------------

B. Sequence Information: Starting position 1778 accession M34058
Exon2     GGGTTTCTGATAGGCACTGACTCTCTCTGCCTATTGGTCTATTTTCCCACCCTTAGGCTGCT
HBBIIIF   ---------------------------------------------------------------------------------------------------
HBB1F     ----------------------------------------------------------CTATTTTCCCACCCTTAGGCTGCT
HBBIIIAR  ----------------------------------------TCTGCCTATTGGTCTATTTTCCCACCCTTAGGCTGCT

Exon2     GGTGGTCTACCCTTGGACCCAGAGGTTCTTTGAGTCCTTTGGGGATCTGTCCACTCCTGATG
HBBIIIF   ----------------------------------------------------GGGGATCTGTCCACTCCTGATG
HBB1F     GGTGGTCTACCCTTGGACCCAGAGGTTCTTTGAGTCCTTTGGGGATCTGTCCACTCCTGATG
HBBIIIAR  GGTGGTCTACCCTTGGACCCAGAGGTTCTTTGAGTCCTTTGGGGATCTGTCCACTCCTGATG

Exon2     CTGTTATGGGCAACCCTAAGGTGAAGGCTCATGGCAAGAAAGTGCTCGGTGCCTTTAGTG
HBBIIIF   CTGTTATGGGCAACCCTAAGGTGAAGGCTCATGGCAAGAAAGTGCTCGGTGCCTTTAGTG
HBB1F     CTGTTATGGGCAACCCTAAGGTGAAGGCTCATGGCAAGAAAGTGCTCGGTGCCTTTAGTG
HBBIIIAR  CTGTTATGGGCAACCCTAAGGTGAAGGCTCATGGCAAGAAAGTGCTCGGTGCCTTTAGTG

Exon2     ATGGCCTGGCTCACCTGGACAACCTCAAGGGCACCTTTGCCACACTGAGTGAGCTGCACTG
HBBIIIF   ATGGCCTGGCTCACCTGGACAACCTCAAGGGCACCTTTGCCACACTGAGTGAGCTGCACTG
HBB1F     ATGGCCTGGCTCACCTGGACAACCTCAAGGGCACCTTTGCCACACTGAGTGAGCTGCACTG
HBBIIIAR  ATGGCCTGGCTCACCTGGACAACCTCAAGGGCACCTTTGCCACACTGAGTGAGCTGCACTG

Exon2     TGACAAGCTGCACGTGGATCCTGAGAACTTCAGGGTGAGTCTATGGGACCCTTGATGTTTT
```

HBBIIIF TGACAAGCTGCACGTGGATCCTGAGAACTTCAGGGTGAGTCTATGGGACCCTTGATGTTTT
HBB1F TGACAAGCTGCACGTGGATCCTGAGAACTTCAGGGTGAGTCTATGGGACCCTTGATGTTTT
HBBIIIAR TGACAAGCTGCACGTGGATCCTGAGAA-
Exon2 CTTTCCCCTTCTTTTCTATGGTTAAGTTCATGTCATAGGAAGGGGAGAAGTAACAGGGTAC
HBBHIF CTTTCCCCTTCTTTTCTATGGTTAAGTTCATGTCATAGGAAGGGGAGAAGTAACAGGGTAC
HBB1F CTTTCCCCTTCTTTTCTATGGTTAAGTTCATGTCATAGGAAGGGGAGAAGTAACAGGGTAC
HBBIIIAR ----------

Exon2 AGTTTAGAATGGGAAACAGACGAATGATTGCATCAGTGTGGAAGTCTCAGGATCGTTTTA
HBBIIIF AGTTTAGAATGGGAAACAGACGAATGATTGCATCAGTGTGGAAGTCTCAGGAT----------
HBB1F AGTTTAGA----------
HBBIIIAR ----------

Exon2 GTTTCTTTTATTTGCT----------
HBBIIIF ----------
HBB1F ----------
HBBIIIAR ----------

REFERENCES

Alberts B BD, Lewis J, Raff M, Roberts K, Watson JD (1994) Molecular Biology of the Cell. New York.

Alberts BD, Lewis J, Raff M, Roberts K, Watson JD (1994) Molecular Biology of the Cell. New York.

Angel JL (1966) Porotic hyperostosis, anemias, malarias, and marshes in the prehistoric Eastern Mediterranean. Science *153:*760-3.

Aula N, and Aula P (2006) Prenatal diagnosis of free sialic acid storage disorders (SASD). Prenat Diagn *26:*655-8.

Beaulne C (2004) The Nature and Structure of DNA Through the Process of Degradation. Masters, Lakehead University, Thunder Bay.

Bickley (2007) Bates' Guide to Physical Examination and History Taking. Philadelphia: Lippincott Williams & Wilkins.

Brandis JW (1999) Dye structure affects Taq DNA polymerase terminator selectivity. Nucleic Acids Res *27:*1912-8.

Brookes AJ (1999) The essence of SNPs. Gene *234:*177-86.

Brownie J, Shawcross S, Theaker J, Whitcombe D, Ferrie R, Newton C, and Little S (1997) The elimination of primer-dimer accumulation in PCR. Nucleic Acids Res *25:*3235-41.

Buchanan GR, DeBaun MR, Quinn CT, and Steinberg MH (2004) Sickle cell disease. Hematology (Am Soc Hematol Educ Program):35-47.

Caramelli D, Lalueza-Fox C, Vernesi C, Lari M, Casoli A, Mallegni F, Chiarelli B, Dupanloup I, Bertranpetit J, Barbujani G, and Bertorelle G (2003) Evidence for a genetic discontinuity between Neandertals and 24,000-year-old anatomically modern Europeans. Proc Natl Acad Sci U S A *100:*6593-7.

Carlson CS, Newman TL, and Nickerson DA (2001) SNPing in the human genome. Curr Opin Chem Biol *5:*78-85.

Chamberlain JS, Gibbs RA, Ranier JE, Nguyen PN, and Caskey CT (1988) Deletion screening of the Duchenne muscular dystrophy locus via multiplex DNA amplification. Nucleic Acids Res *16:*11141-56.

Chen X, and Sullivan PF (2003) Single nucleotide polymorphism genotyping: biochemistry, protocol, cost and throughput. Pharmacogenomics J *3:*77-96.

Chou Q, Russell M, Birch DE, Raymond J, and Bloch W (1992) Prevention of pre-PCR mis-priming and primer dimerization improves low-copy-number amplifications. Nucleic Acids Res *20:*1717-23.

Cipollaro M, Galderisi U, and Di Bernardo G (2005) Ancient DNA as a multidisciplinary experience. J Cell Physiol *202:*315-22.

Collins FS, Weissman, S. M. (1984) The Molecular Genetics of Human Hemoglobin: Progress in Nucleic Acid Resaerch and Molecular Biology, pp. 315-462.

Cooper A, Lalueza-Fox C, Anderson S, Rambaut A, Austin J, and Ward R (2001) Complete mitochondrial genome sequences of two extinct moas clarify ratite evolution. Nature *409:*704-7.

Deragon JM, Sinnett D, Mitchell G, Potier M, and Labuda D (1990) Use of gamma irradiation to eliminate DNA contamination for PCR. Nucleic Acids Res *18:*6149.

Dieffenbach CW, Lowe TM, and Dveksler GS (1993) General concepts for PCR primer design. PCR Methods Appl *3:*S30-7.
Eaton WA, Henry ER, Hofrichter J, Bettati S, Viappiani C, and Mozzarelli A (2007) Evolution of allosteric models for hemoglobin. IUBMB Life *59:*586-99.
Erlich HA (1989) PCR technology : principles and applications for DNA amplification. New York, NY: Stockton Press.
Faerman M, Filon D, Kahila G, Greenblatt CL, Smith P, and Oppenheim A (1995) Sex identification of archaeological human remains based on amplification of the X and Y amelogenin alleles. Gene *167:*327-32.
Faerman M, Nebel A, Filon D, Thomas MG, Bradman N, Ragsdale BD, Schultz M, and Oppenheim A (2000) From a dry bone to a genetic portrait: a case study of sickle cell anemia. Am J Phys Anthropol *111:*153-63.
Filon D, Faerman M, Smith P, and Oppenheim A (1995) Sequence analysis reveals a beta-thalassaemia mutation in the DNA of skeletal remains from the archaeological site of Akhziv, Israel. Nat Genet *9:*365-8.
Fucharoen S, Changtrakun Y, Surapot S, Fucharoen G, and Sanchaisuriya K (2002) Molecular characterization of Hb D-Punjab [beta121(GH4)Glu-->Gln] in Thailand. Hemoglobin *26:*261-9.
Gernaey AM, Minnikin DE, Copley MS, Dixon RA, Middleton JC, and Roberts CA (2001) Mycolic acids and ancient DNA confirm an osteological diagnosis of tuberculosis. Tuberculosis (Edinb) *81:*259-65.
Gilbert MT, Bandelt HJ, Hofreiter M, and Barnes I (2005) Assessing ancient DNA studies. Trends Ecol Evol *20:*541-4.
Graur D, and Pupko T (2001) The Permian bacterium that isn't. Mol Biol Evol *18:*1143-6.
Gu J (1995) In situ PCR and related technology. Cambridge, MA, USA: Birkhèauser Boston.
Guillaudeux T, Janer M, Wong GK, Spies T, and Geraghty DE (1998) The complete genomic sequence of 424,015 bp at the centromeric end of the HLA class I region: gene content and polymorphism. Proc Natl Acad Sci U S A *95:*9494-9.
Hagelberg E, Bell LS, Allen T, Boyde A, Jones SJ, and Clegg JB (1991) Analysis of ancient bone DNA: techniques and applications. Philos Trans R Soc Lond B Biol Sci *333:*399-407.
Handt O, Hoss M, Krings M, and Paabo S (1994) Ancient DNA: methodological challenges. Experientia *50:*524-9.
Handt O, Krings M, Ward RH, and Paabo S (1996) The retrieval of ancient human DNA sequences. Am J Hum Genet *59:*368-76.
Hansen JN (1974) Isolation of higher molecular weight DNA from Bacillus Cereus T using proteinase K. Prep Biochem *4:*473-88.
Harris E, Kropp G, Belli A, Rodriguez B, and Agabian N (1998) Single-step multiplex PCR assay for characterization of New World Leishmania complexes. J Clin Microbiol *36:*1989-95.
Harris S, and Jones DB (1997) Optimisation of the polymerase chain reaction. Br J Biomed Sci *54:*166-73.
Hartl D, and Jones, E. (2005) Genetics: Analysis of Genes and Genomes: Jones & Barlett.
Hauswirth WW (1994) Ancient DNA: an introduction. Experientia *50:*521-3.
Hayter S (2007) Improving Genetic Profiling Techniques for Low Copy Number DNA, Lakehead, Thunder Bay

Hendolin PH, Markkanen A, Ylikoski J, and Wahlfors JJ (1997) Use of multiplex PCR for simultaneous detection of four bacterial species in middle ear effusions. J Clin Microbiol *35:*2854-8.

Henegariu O, Heerema NA, Dlouhy SR, Vance GH, and Vogt PH (1997) Multiplex PCR: critical parameters and step-by-step protocol. Biotechniques *23:*504-11.

Henegariu O, Hirschmann P, Kilian K, Kirsch S, Lengauer C, Maiwald R, Mielke K, and Vogt P (1994) Rapid screening of the Y chromosome in idiopathic sterile men, diagnostic for deletions in AZF, a genetic Y factor expressed during spermatogenesis. Andrologia *26:*97-106.

Heredia A, Soriano V, Weiss SH, Bravo R, Vallejo A, Denny TN, Epstein JS, and Hewlett IK (1996) Development of a multiplex PCR assay for the simultaneous detection and discrimination of HIV-1, HIV-2, HTLV-I and HTLV-II. Clin Diagn Virol *7:*85-92.

Herrmann B, and Hummel S (1994) General Aspects of Sample Preparation. In B Herrmann and S Hummel (eds.): Ancient DNA. New York: Springer-Verlag, pp. 59-68.

Hershkovitz I, Rothschild BM, Latimer B, Dutour O, Leonetti G, Greenwald CM, Rothschild C, and Jellema LM (1997) Recognition of sickle cell anemia in skeletal remains of children. Am J Phys Anthropol *104:*213-26.

Higuchi RG, Wrischnik LA, Oakes E, George M, Tong B, and Wilson AC (1987) Mitochondrial DNA of the extinct quagga: relatedness and extent of postmortem change. J Mol Evol *25:*283-7.

Hofreiter M, Serre D, Poinar HN, Kuch M, and Paabo S (2001) Ancient DNA. Nat Rev Genet *2:*353-9.

Horton R, Niblett D, Milne S, Palmer S, Tubby B, Trowsdale J, and Beck S (1998) Large-scale sequence comparisons reveal unusually high levels of variation in the HLA-DQB1 locus in the class II region of the human MHC. J Mol Biol *282:*71-97.

Hoss M, Jaruga P, Zastawny TH, Dizdaroglu M, and Paabo S (1996) DNA damage and DNA sequence retrieval from ancient tissues. Nucleic Acids Res *24:*1304-7.

Huehns ER (1974) The Structure and Function of Heamoglobin: Clinical Disorders Due to Abnormal Haemoglobin Structure. Oxford: Blackwell.

Hummel S (2003) Ancient DNA typing : methods, strategies, and applications. New York: Springer.

Innis MA, Myambo KB, Gelfand DH, and Brow MA (1988) DNA sequencing with Thermus aquaticus DNA polymerase and direct sequencing of polymerase chain reaction-amplified DNA. Proc Natl Acad Sci U S A *85:*9436-40.

Kemp BM, and Smith DG (2005) Use of bleach to eliminate contaminating DNA from the surface of bones and teeth. Forensic Sci Int *154:*53-61.

Kirk BW, Feinsod M, Favis R, Kliman RM, and Barany F (2002) Single nucleotide polymorphism seeking long term association with complex disease. Nucleic Acids Res *30:*3295-311.

Kruglyak L (1999) Prospects for whole-genome linkage disequilibrium mapping of common disease genes. Nat Genet *22:*139-44.

Kwok PY (2001) Methods for genotyping single nucleotide polymorphisms. Annu Rev Genomics Hum Genet *2:*235-58.

Kwok PY, and Chen X (2003) Detection of single nucleotide polymorphisms. Curr Issues Mol Biol *5:*43-60.

Kwok S, and Higuchi R (1989) Avoiding false positives with PCR. Nature *339:*237-8.

Laberge M, and Yonetani T (2007) Common dynamics of globin family proteins. IUBMB Life *59:*528-34.
Landegren U, Nilsson M, and Kwok PY (1998) Reading bits of genetic information: methods for single-nucleotide polymorphism analysis. Genome Res *8:*769-76.
Larcombe LA, Nickerson P, Hoppa RD, and Matheson C (2005) Detection of a single nucleotide polymorphism in the IL-6 promoter region of ancient nuclear DNA. Infect Genet Evol *5:*117-22.
Li WH, and Sadler LA (1991) Low nucleotide diversity in man. Genetics *129:*513-23.
Li Y, Mitaxov V, and Waksman G (1999) Structure-based design of Taq DNA polymerases with improved properties of dideoxynucleotide incorporation. Proc Natl Acad Sci U S A *96:*9491-6.
Lindahl T (1993) Instability and decay of the primary structure of DNA. Nature *362:*709-15.
Lindahl T, and Andersson A (1972) Rate of chain breakage at apurinic sites in double-stranded deoxyribonucleic acid. Biochemistry *11:*3618-23.
Lindahl T, and Nyberg B (1972) Rate of depurination of native deoxyribonucleic acid. Biochemistry *11:*3610-8.
Livingstone FB (1895) Frequency of Hemoglobin Variants. New York.
Luftig M RS (2001) DNA and forensic science. New England Law Review *35:*609-13.
Mansfield ES, Robertson JM, Lebo RV, Lucero MY, Mayrand PE, Rappaport E, Parrella T, Sartore M, Surrey S, and Fortina P (1993) Duchenne/Becker muscular dystrophy carrier detection using quantitative PCR and fluorescence-based strategies. Am J Med Genet *48:*200-8.
Markoulatos P, Samara V, Siafakas N, Plakokefalos E, Spyrou N, and Moncany ML (1999) Development of a quadriplex polymerase chain reaction for human cytomegalovirus detection. J Clin Lab Anal *13:*99-105.
Markoulatos P, Siafakas N, Katsorchis T, and Moncany M (2003) Multiplex PCR: rapid DNA cycling in a conventional thermal cycler. J Clin Lab Anal *17:*108-12.
Markoulatos P, Siafakas N, and Moncany M (2002) Multiplex polymerase chain reaction: a practical approach. J Clin Lab Anal *16:*47-51.
McPherson MJ, Moller, S. G. (2000) *Pcr*. New York: Springer-Verlag.
Miller RG, Segal JB, Ashar BH, Leung S, Ahmed S, Siddique S, Rice T, and Lanzkron S (2006) High prevalence and correlates of low bone mineral density in young adults with sickle cell disease. Am J Hematol *81:*236-41.
Mitchell D, Willerslev E, and Hansen A (2005) Damage and repair of ancient DNA. Mutat Res *571:*265-76.
Mullis K, Faloona F, Scharf S, Saiki R, Horn G, and Erlich H (1986) Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. Cold Spring Harb Symp Quant Biol *51 Pt 1:*263-73.
Mullis KB (1990) Target amplification for DNA analysis by the polymerase chain reaction. Ann Biol Clin (Paris) *48:*579-82.
Mutirangura A, Greenberg F, Butler MG, Malcolm S, Nicholls RD, Chakravarti A, and Ledbetter DH (1993) Multiplex PCR of three dinucleotide repeats in the Prader-Willi/Angelman critical region (15q11-q13): molecular diagnosis and mechanism of uniparental disomy. Hum Mol Genet *2:*143-51.
Mutter GL, and Boynton KA (1995) PCR bias in amplification of androgen receptor alleles, a trinucleotide repeat marker used in clonality studies. Nucleic Acids Res *23:*1411-8.

Nachman MW, Bauer VL, Crowell SL, and Aquadro CF (1998) DNA variability and recombination rates at X-linked loci in humans. Genetics *150:*1133-41.

Paabo S (1989) Ancient DNA: extraction, characterization, molecular cloning, and enzymatic amplification. Proc Natl Acad Sci U S A *86:*1939-43.

Paabo S, Poinar H, Serre D, Jaenicke-Despres V, Hebler J, Rohland N, Kuch M, Krause J, Vigilant L, and Hofreiter M (2004) Genetic analyses from ancient DNA. Annu Rev Genet *38:*645-79.

Papathanasiou A (2005) Health status of the Neolithic population of Alepotrypa Cave, Greece. Am J Phys Anthropol *126:*377-90.

Poinar H (2003) Nuclear Gene Sequences from a Late Pleistocene Sloth Coprolite. Current Biology *13:*1150-2.

Rees DC, Styles L, Vichinsky EP, Clegg JB, and Weatherall DJ (1998) The hemoglobin E syndromes. Ann N Y Acad Sci *850:*334-43.

Repp R, Rhiel S, Heermann KH, Schaefer S, Keller C, Ndumbe P, Lampert F, and Gerlich WH (1993) Genotyping by multiplex polymerase chain reaction for detection of endemic hepatitis B virus transmission. J Clin Microbiol *31:*1095-102.

Risch N, and Merikangas K (1996) The future of genetic studies of complex human diseases. Science *273:*1516-7.

Rithidech KN, Dunn JJ, and Gordon CR (1997) Combining multiplex and touchdown PCR to screen murine microsatellite polymorphisms. Biotechniques *23:*36, 40, 42, 44.

Rohland N, and Hofreiter M (2007) Ancient DNA extraction from bones and teeth. Nat Protoc *2:*1756-62.

Rosenblum BB, Lee LG, Spurgeon SL, Khan SH, Menchen SM, Heiner CR, and Chen SM (1997) New dye-labeled terminators for improved DNA sequencing patterns. Nucleic Acids Res *25:*4500-4.

Ruano G, Fenton W, and Kidd KK (1989) Biphasic amplification of very dilute DNA samples via 'booster' PCR. Nucleic Acids Res *17:*5407.

Rys PN, and Persing DH (1993) Preventing false positives: quantitative evaluation of three protocols for inactivation of polymerase chain reaction amplification products. J Clin Microbiol *31:*2356-60.

Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, Sherry S, Mullikin JC, Mortimore BJ, Willey DL, Hunt SE, Cole CG, Coggill PC, Rice CM, Ning Z, Rogers J, Bentley DR, Kwok PY, Mardis ER, Yeh RT, Schultz B, Cook L, Davenport R, Dante M, Fulton L, Hillier L, Waterston RH, McPherson JD, Gilman B, Schaffner S, Van Etten WJ, Reich D, Higgins J, Daly MJ, Blumenstiel B, Baldwin J, Stange-Thomann N, Zody MC, Linton L, Lander ES, and Altshuler D (2001) A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature *409:*928-33.

Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, and Erlich HA (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science *239:*487-91.

Sanger F, Nicklen S, and Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A *74:*5463-7.

Schork NJ, Cardon LR, and Xu X (1998) The future of genetic epidemiology. Trends Genet *14:*266-72.

Shuber AP, Skoletsky J, Stern R, and Handelin BL (1993) Efficient 12-mutation testing in the CFTR gene: a general model for complex mutation analysis. Hum Mol Genet *2:*153-8.

Steinberg MH (1998) Pathophysiology of sickle cell disease. Baillieres Clin Haematol *11:*163-84.

Stewart JE, Aagaard PJ, Pokorak EG, Polanskey D, and Budowle B (2003) Evaluation of a multicapillary electrophoresis instrument for mitochondrial DNA typing. J Forensic Sci *48:*571-80.

Syvanen AC (2001) Accessing genetic variation: genotyping single nucleotide polymorphisms. Nat Rev Genet *2:*930-42.

Tindall KR, and Kunkel TA (1988) Fidelity of DNA synthesis by the Thermus aquaticus DNA polymerase. Biochemistry *27:*6008-13.

Vandenvelde C, Verstraete M, and Van Beers D (1990) Fast Multiplex polymerase chain reaction on boiled clinical samples for rapid viral diagnosis. J Virol Methods *30:*215-27.

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M, Subramanian G, Thomas PD, Zhang J, Gabor Miklos GL, Nelson C, Broder S, Clark AG, Nadeau J, McKusick VA, Zinder N, Levine AJ, Roberts RJ, Simon M, Slayman C, Hunkapiller M, Bolanos R, Delcher A, Dew I, Fasulo D, Flanigan M, Florea L, Halpern A, Hannenhalli S, Kravitz S, Levy S, Mobarry C, Reinert K, Remington K, Abu-Threideh J, Beasley E, Biddick K, Bonazzi V, Brandon R, Cargill M, Chandramouliswaran I, Charlab R, Chaturvedi K, Deng Z, Di Francesco V, Dunn P, Eilbeck K, Evangelista C, Gabrielian AE, Gan W, Ge W, Gong F, Gu Z, Guan P, Heiman TJ, Higgins ME, Ji RR, Ke Z, Ketchum KA, Lai Z, Lei Y, Li Z, Li J, Liang Y, Lin X, Lu F, Merkulov GV, Milshina N, Moore HM, Naik AK, Narayan VA, Neelam B, Nusskern D, Rusch DB, Salzberg S, Shao W, Shue B, Sun J, Wang Z, Wang A, Wang X, Wang J, Wei M, Wides R, Xiao C, Yan C, et al. (2001) The sequence of the human genome. Science *291:*1304-51.

Wagner A, et al. (1994) *Surveys of Gene Families Using Polymerase Chain Reaction: PCR Selection and PCR Drift.* Systematic Biology *43:*250-261.

Wang DG, Fan JB, Siao CJ, Berno A, Young P, Sapolsky R, Ghandour G, Perkins N, Winchester E, Spencer J, Kruglyak L, Stein L, Hsie L, Topaloglou T, Hubbell E, Robinson E, Mittmann M, Morris MS, Shen N, Kilburn D, Rioux J, Nusbaum C, Rozen S, Hudson TJ, Lipshutz R, Chee M, and Lander ES (1998) Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. Science *280:*1077-82.

Wang WY, and Todd JA (2003) The usefulness of different density SNP maps for disease association studies of common variants. Hum Mol Genet *12:*3145-9.

Watson JD, and Crick FH (1953a) Genetical implications of the structure of deoxyribonucleic acid. Nature *171:*964-7.

Watson JD, and Crick FH (1953b) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature *171:*737-8.

Weatherall DJ, and Clegg JB (1976) Molecular Genetics of Human Hemoglobin. Annu Rev Genet *10:*157-78.

Weatherall DJ, and Clegg JB (1976) Molecular genetics of human hemoglobin. Annu Rev Genet *10:*157-78.

Weatherall DJ, and Clegg JB (2001) Inherited haemoglobin disorders: an increasing global health problem. Bull World Health Organ *79:*704-12.

WHO (1989) Guidelines for the control of haemoglobin disorders. Report on the VIth Annual Meeting of the World Health Organization Working Group on Haemoglobinopathies.
Willerslev E, and Cooper A (2005) Ancient DNA. Proc Biol Sci *272:*3-16.
Xu B, Tubbs RR, and Kottke-Marchant K (2005) Molecular genetic testing of polymorphisms associated with venous thrombosis: a review of molecular technologies. Diagn Mol Pathol *14:*193-202.
Zhao R (2005) From single cell gene-based diagnostics to diagnostic genomics: current applications and future perspectives. Clin Lab Sci *18:*254-62.
Zhu YS, Isaacs ST, Cimino CD, and Hearst JE (1991) The use of exonuclease III for polymerase chain reaction sterilization. Nucleic Acids Res *19:*2511.
Zimmermann K, Schogl D, Plaimauer B, and Mannhalter JW (1996) Quantitative multiple competitive PCR of HIV-1 DNA in a single reaction tube. Biotechniques *21:*480-4.